Dynal Patel

Photo searching on small screen devices

Thesis Presented for the Degree of

DOCTOR OF PHILOSOPHY

in the Department of Computer Science

UNIVERSITY OF CAPE TOWN

February 2007

Supervisors

Gary Marsden

Matt Jones

Steve Jones

Abstract

The aim of this thesis is to improve HCI (Human-Computer Interaction) knowledge in the design of the next generation of photo search tools on small display devices. Today, these devices have all the ingredients for a truly mobile photo collection, such as large storage, multiple networking capabilities and high resolution screens. However, they lack the tools for searching through large collections of photographs. This is particularly important as users have expressed a desire to store images on mobile devices in the long term.

No substantial research has looked at addressing users searching needs. Few researchers have considered the importance of supporting both searching and browsing to cater for user needs. None that we could find have assessed the potential impact of adding desktop-based annotation capabilities. Consequently, this thesis seeks to address these challenges and provide an empirical foundation for the design of photo search tools.

To achieve these objectives an iterative user-centered design methodology was employed. The end practical result was a single photo search interface that incorporates the best traits of a variety of tools to support search. The thesis reflects on each cycle in the iterative design process. The first major area of contribution to the field of HCI improves existing knowledge on photo searching behavior by providing a number of empirically grounded findings about searching behavior. It identifies some of the core factors that influence search strategies and outlines a conceptual framework to guide the design of future systems. The second area of contribution is a single photo searching tool for small display devices that is based on iterative studies of various user interface designs. It integrates multiple search methods within a single user interface. In contrast to previous research in this area, the design is centered on locating events rather than individual images as we found that people naturally associate photographs with events when searching.

Declaration

Some aspects of the work described in this thesis were presented in the following publications:

1. Patel, D., Marsden, G., Jones, M., and Jones, S. (2006, September). Improving photo searching interfaces for small-screen mobile computers. *In Proceedings of the 8th Conference on Human-Computer interaction with Mobile Devices and Service* (MobileHCI '06s), vol. 159, in Helsinki, Finland, ACM Press, New York, NY, pp. 149-156.

2. Patel, D., and Marsden, G. (2006). Investigating the use of photo collection structures for photo searching, *In Proceeding Chi-SA 2006*, ACM Press, in Cape Town, South Africa, pp. 41-50.

3. Jones, S., Jones, M., Marsden, G., Patel, D., and Cockburn, A. (2005, September). An evaluation of integrated zooming and scrolling on small screens. *International Journal of Human.-Computer Studies*. 63 (3), pp. 271-303.

4. Patel, D., Marsden, G., Jones, S., and Jones, M. (2004, SeptemberAn evaluation of techniques for browsing photograph collections on small displays. *In Proceedings Mobile HCI*, in Glasgow, Scotland, pages 132-143.

As the first author in papers 1, 2 and 4, I was responsible for designing, implementing and evaluating all the prototypes. I was also responsible for writing the papers. My supervisors provided guidance in refining the papers for submission. As a co-author in paper 3, I contributed significantly in the design, development and evaluation. I developed one of two software prototypes. I was also involved in the experimental design and also developed some logging software to automatically capture data during the experiment. Steve Jones conducted the experiment and wrote the paper. As one of the co-authors, I was also involved in reviewing the paper before and after submission.

Acknowledgements

Firstly, I would like to thank my main supervisor (Gary Marsden) for giving me the freedom to study whatever I was interested in, for keeping me motivated, and steering successfully through the many challenges (always going beyond the call of duty I must add). Special thanks to Gary and Gil for the generosity they have shown me throughout my studies.

I am so grateful for the advice and encouragement from my co-supervisors Matt and Steve. Steve especially for making sure that everything was done rigorously and Matt for being a never ending source of ideas. It was really fun working with you both.

I would also like to thank Michael Smith, Richard Harper and George Buchanan for their feedback towards the end of the thesis. Many other people in the HCI community also played a part (in no particular order): Steve Brewster, Vincent Buil, Ben Shneiderman, Jenny Preece, Ian Witten, David Nichols, Edwin Blake, Bjorn, Malin, all the members of the CVC lab past and present and also to the reviewers and examiners. I am also really grateful to all the participants that took part in the experiments.

I would also like to acknowledge the following people for sponsoring my undergraduate studies, without whom I would have never got this far: Bentleys, Haighs, Van der Maas, Groves, Whites, Reas, Fishers, Poidevins, Kayes, Sucklings, Wyatts, Jacks, Hillary Millard, Wagers, and Patchings. For those of you who chose to remain anonymous and anyone I might have forgotten a special thank you to you too. I would also like to thank the Beit Trust and the Harry Crossley foundation for funding my postgraduate study (Masters Degree and the PhD when I was upgraded). Thank you for funding the conference trips and providing the research materials when I needed them.

I would also like to thank my friends and family for showing me that there is more to life than a PhD. Special thanks to Andrew and Cilla for their support and guidance. Hugs to Lyn and Raja for being so thoughtful and kind. Thank you both for the endless messages of encouragement.

Lastly, I would like to dedicate this thesis to my mother, Doreen Chitambala Patel. Thank you for the countless hours of prayer. The completion of this thesis marks the answer to that prayer.

Contents

List of figures

List of tables

Chapter 1

Introduction

1.1 Introduction

In the early 1800's, photography was the business and pastime of a small number of individuals, experts in both the technology for capturing images and the chemistry of processing them. Photographs would depict the family and its hierarchy [42]. They were seen as status symbols, depicting power and wealth [67]. Portrait studios would use all kinds of props to create a setting, a wished for situation perhaps or even a fictitious past [11].

(a) (b)

Figure 1.1: The Kodak Brownie (a) and Nokia N90 (b)

The introduction of the inexpensive Kodak brownie (see Figure 1.1a) in the early 1900's sparked a major change in the nature of photography by taking photography out of the studio and making it more accessible to the masses. The Kodak Eastman company proposed a solution where they did all the processing (developing and printing), and all the user had to do was take the picture. Unfortunately, the premium attached to this service kept the price of photography relatively high [67].

The introduction of the first Polaroid Camera in 1947 brought down some of the costs by cutting the professionals out of the loop. However, the key factor in reducing the costs was the transition from film to digital photography which began in the early 1970's [67]. Beyond the initial expenditure, the costs of film and printing no longer applied [62]. This alleviated the sense of scarcity associated with film-based photography, enabling users to take significantly more photographs of their everyday lives [68]. Instead of traditional formal pictures, they showed 'life as action, interaction with informal dress' [42]. The ability to instantly view and delete photographs led to people taking more shots than usual and only the best ones were kept [68]. The transition to a digital medium also made it easier to edit, store, send and share photographs.

The recent convergence between mobile phones and digital cameras has sparked yet another change (see Figure 1.1b). The 'always in hand' capability of such devices has meant that photographs can now to be taken anytime and anywhere, resulting in photo collections growing more rapidly than ever and enabling people to amass large collections in relatively short time periods. Although these devices have ever increasing capacities to store photographs, their use presents users with a challenge, as the screens on which these images are viewed and browsed is relatively small. A question that arises then, is how may a user be supported in searching for photographs on such a device with a limited display space?

1.2 Photo search solutions

Two approaches for locating images are browsing and searching. When users have a predefined goal or criteria in mind, a search is often used to find a particular image. When there is no conscious requirement, users often elect to browse through the collection, recognizing things as they see them. In fact, the two approaches are quite complementary [106]. There are times when a user might elect to browse through a collection to satisfy a requirement, as an alternative to issuing a query. For example, if the requirements are difficult to express. Similarly, when browsing, it is possible for the requirements to become more defined as new information is discovered. Depending on the task at hand, browsing can take on a range of different functions from specific to purposeful to serendipitous [66].

Most photo search tools that support query-based searching rely on manually annotated metadata for each image [57][113][135]. With personal photography, people are reluctant to annotate the bulk of their photo collection [62][105]. The time and effort required to annotate each photograph far out-weighs any perceived benefits [105]. As a result, researchers have relied on methods that distinguish photographs based on the image content, as this data is implicitly available. These content-based methods have been quite successful in locating photographs that have similar color compositions. However, they are unable to identify and interpret meaningful features in images. Consequently, they have not been particularly useful, as people search for images based on semantic properties rather than visual properties [62]. The 'meaningful identification and interpretation of features in photographs has been described as an AI-complete problem which requires solving strong AI in general' [106].

One solution is to rely on humans to do this complex computation by exploiting the human ability to process visual information rapidly. Many commercial image browsing tools make use of this approach by maximizing the use of the screen real-estate, so that at any given time as many thumbnails are shown as possible. This technique works particularly well on high resolution screens where a large number of thumbnails can be displayed. However, the technique is not as effective on handhelds, where the lower resolution screens only allow a few images to be shown at a time. Given that about half the cerebral cortex is dedicated to image processing alone [136], we believe it is still worth investigating other techniques that may exploit the human ability to rapidly interpret large amounts of visual data.

1.3 Limitations of previous research

When searching through old photographs, people are most likely to look for: an *event* (set of photographs relating to a particular well defined event); a *single* (individual remembered photograph), and a *property* (sets of photographs with a common theme) [105]. Unfortunately, the majority of studies in this area have not considered supporting users in performing these tasks [62]. Most research studies have tended to focus on just one (locating a *single*) [29][65][128][118] or two (locating a *single* and *property*) [44] of these tasks. This raises questions about their ecological validity, particularly as studies by Rodden and Wood [105] have found that the most common photo search task is locating events.

Similarly, very few researchers have considered the importance of supporting both searching and browsing. It is unlikely that any single approach is suitable in all circumstances. For example, a search is likely to be most effective when user needs are well defined. However, as user needs become more vague and difficult to express, an exploratory approach that exposes the user to the information is likely to be more appropriate. The ideal approach for supporting user needs is likely to be a more inclusive one that accommodates both methods of information access.

Very little photo organization and structuring is performed on handheld devices. Most photo management is performed on desktop computers using specialist software such as iPhoto [4]. Photo collection structures are then migrated to the handheld device during a synchronization process. Devices such as the iPhone allow a mirror copy of an entire photo collection to be placed on the device. Replicating a photo collection's structure on a handheld has a number of advantages: it provides a structure that is already familiar; it allows the photo collection to be accessed anytime, anywhere; and it also provides a way of backing up the photo collection. Wilhelm and her colleagues [131] in their investigation of photo annotation on camera phones concluded that there is a 'need to develop hybrid solutions that integrate desktop and mobile application components into more complete and appropriate solutions.' They recommend adding desktop-based annotation capabilities to photo search tools on mobile devices. We were unable to find any research that has looked at how user-inputted metadata from the desktop might be used to support photo search on a handheld device.

1.4 Research Questions

The main aim of this thesis is to provide an empirical foundation for the design of photo search tools on small display devices, taking into account the challenges (or research limitations) discussed above. The key objectives are:

1. To devise a set of empirically grounded guidelines for designing photo search interfaces on small display devices.

2. To develop a photo search tool that addresses the limitations of current systems on small display devices. In contrast to previous work, the design of search tool must take into consideration the task types (*events*, *singles* and *properties*), the methods of information access (searching and browsing) and current annotation practices.

1.5 Justification for the research

The need for adequate photo search solutions on handhelds has been advocated in a number of research papers [29][44][61]. The arguments have been wide ranging. A techno-centric argument has focused on the fact that mobile devices have all the necessary ingredients (large storage, multiple networking capabilities, and high resolution screens) for a truly mobile collection. For example, the current iPod photo and the iPod phone are capable of storing more than 25,000 images. However, what is lacking are tools that allow users to sift through potentially large numbers of photographs, especially as people have expressed their desire to store images in the long term on their mobile device [61]. Another argument is that as storage increases, users will be less motivated to transfer photos to the PC, so searching activity will naturally migrate from the PC to the mobile device. Kindberg at al. [61] have stressed the need for adequate photo search tools to support the activities around mobile photography, such as storytelling and reminiscing, which typically require users to sift through old photographs. Photo search tools are also needed to support photo sharing on mobile devices, especially between groups of people that mutually experience an event and exchange photographs sometime after the event when they next meet each other.

Photo search tools have the potential to impact civil society. For example, Figure 1.2 shows that there are currently about 600 million camera phone users [50]; if only 10% search for images on their phone, saving 1 second per search saves a total of 166666 hours. Of course this crude estimate ignores a number of factors, such as searching during 'dead time.' However, it illustrates the importance of conducting research in this area.

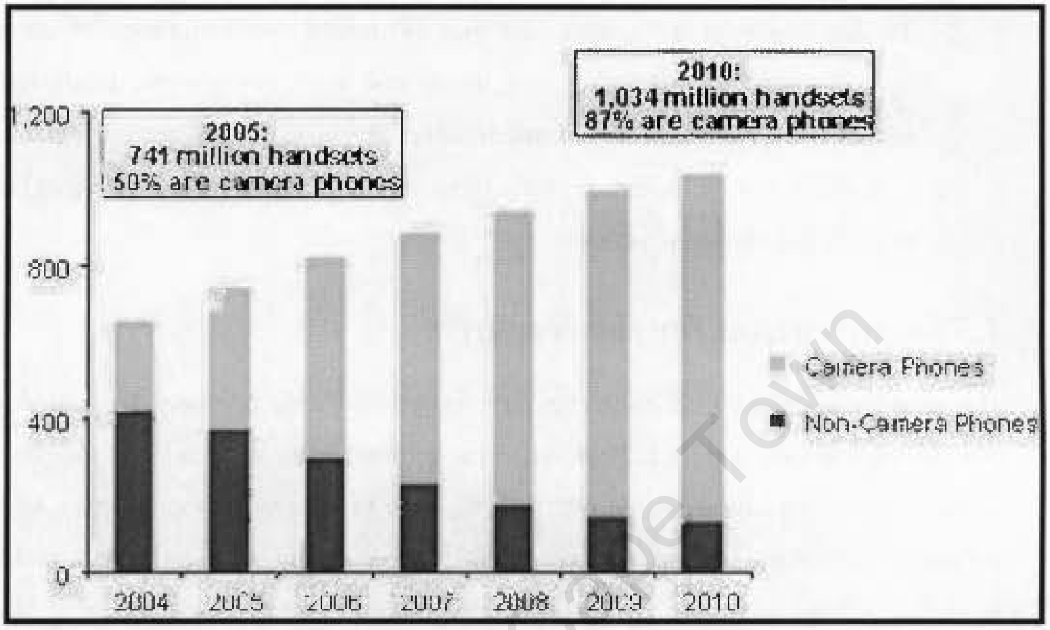
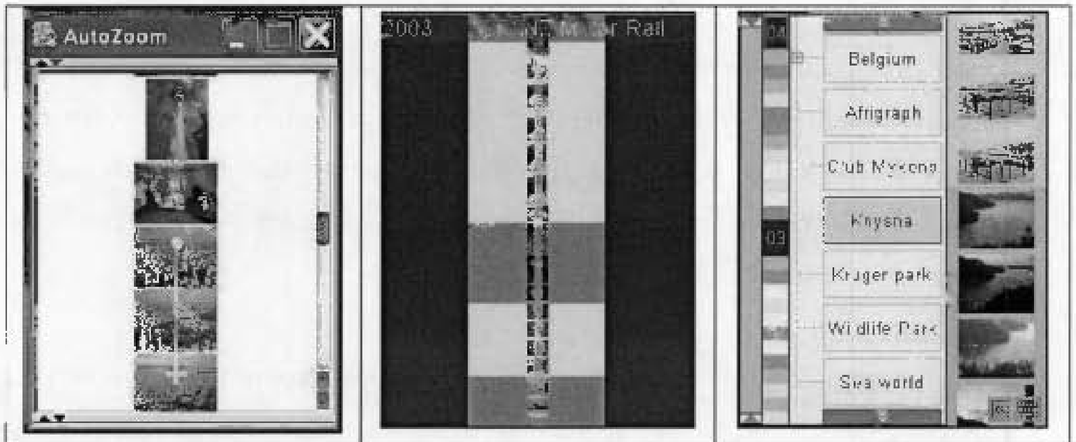


Figure 1.2: Worldwide mobile phone and camera phone shipments [50].

1.6 Methodology

Given the emphasis on supporting user needs, the field of Human-Computer Interaction provides suitable methods for designing, developing and evaluating photo search interfaces. User-centered design is the most widely used methodology (or design philosophy) for developing computer systems that meet user needs. It places users at the center of the design and involves them throughout the design process. It uses an iterative design process for developing, evaluating and refining computer systems. The benefit of using a design based research methodology is that it enables researchers to 'understand the world through the process of gathering requirements and improve the world through the process of design' [16].



Chapter 3	Chapter 4	Chapter 6
A	B	C

Figure 1.3: Three photo search tools that are developed as part of an iterative design process. Figure 1.3a illustrates the initial search tool that supports two *Visual photo search techniques* (see section 1.9 for definition). In Figure 1.3b, the search tool is modified to include contextual cues. In Figure 1.3c, the final search tool integrates multiple search methods which include: a *Keyword search, Timeline browser, Timeline filter, Hierarchical folder-based browser* and two *Visual photo search techniques*.

An iterative design process is used to develop three photo search tools (see Figure 1.3). Figure 1.3a shows the initial search tool that supports two different visual photo search techniques. The requirements for this search tool are based on a user needs analysis in Chapter 2. A *design-first, study-second* approach is used to develop this search tool. Don Norman [89] argues that when user needs have already been investigated and a benchmark exists, a *design-first, study-second* approach is more appropriate than a *study-first, design-second* approach as it avoids replicating existing work. The design is deliberately incremental so that it can be benchmarked against a baseline thumbnail-grid browser. The emphasis here is on improving current photo search tools for small display devices by taking into account users needs when visually searching through photo collections. The next iteration in the cycle is aimed at identifying the problems with the search tool so that they can be addressed to provide a better search experience in terms of supporting users in performing the three common searching tasks. The search tool shown in Figure 1.3b is developed further by adding contextual cues (date and folder information) to the search tool shown in Figure 1.3a. Studies have shown that dates and folder labels are used extensively when searching through a photo collection [62][131]. In the final iteration, the search tool is developed further based on the recommendations in Chapter 4. The search tool (see Figure 1.3c) integrates multiple search methods which include a *Keyword search, Temporal browser, Hierarchical folder-based browser* and two *Visual photo search techniques*.

A set of empirically grounded guidelines are developed for designing future photo search interfaces based on user evaluations of the three search tools. Three common evaluation practices for visual systems [98] are used. A controlled experiment is used to compare the initial photo search tool against the state of the art (see Chapter 3). An observational study (or usability study) is used to assess the suitability of the photo search tool in supporting three common user tasks (see Chapter 4). A further study is conducted in a more realistic setting where the search requirements are not always well-defined. We gauge the utility of the search tool in supporting the three common search tasks as information about a target becomes less well-known (see Chapter 6). By examining these search tools from multiple perspectives, the goal is to gain new insights in to how to support the common search tasks. Both quantitative and qualitative research methods are used to collect data. By using both research methods, the aim is overcome the flaws that are inherent in each method.

1.7 Contributions

The first contribution to field of Human Computer Interaction improves existing knowledge on photo searching behavior by providing a number of empirically grounded findings about searching behavior. One important finding is that searching tasks rely on the ability to locate events rapidly. This research uncovers some of the important factors that impact the way people search for events. One significant factor is the user's memory of an event. A conceptual framework is developed to guide future research and design, based on a thorough investigation of this factor.

The second area of contribution is a single photo search tool for small display devices that is based on iterative studies of various user interface designs that has tried to incorporate the best traits of a variety of search tools to support the common photo search tasks. The tool allows users to perform a query-based search by retrieving events rather than individual photographs. This takes advantage of the fact that people organize photographs by events and provide descriptions for each event. The tool also supports two techniques for visually searching through photographs. The techniques are designed to make optimal use of the human ability to process visual information rapidly. A number of algorithms are proposed to support these techniques in limited environments, tackling three major limitations with mobile devices: low resolution screens, varied input mechanisms, and limited resources such as memory and processing capabilities. A clear implementation guide is provided for developers and practitioners to enable them to replicate these techniques.

1.8 Thesis Outline

Figure 1.4 illustrates the thesis outline, showing the key research components and the types of evaluation in the blue blocks.

Chapter 2 introduces the field of Human-Computer Interaction. It discusses the human factors in photo searching. It describes the linkages between two methods of information access, searching and browsing. It also conducts a critical analysis of previous photo search tools and highlights the limitations of current query-based approaches and visual photo search approaches. It uses framework for navigation to describe some design implications for visual photo search tools.

Chapter 3 describes the development of a photo search tool that is designed to address the limitations of current tools that allow users to visually search through photographs. The search tool supports two techniques that enable users to navigate rapidly through the photo collection. The first technique is an adaptation of a navigation technique called Speed Dependant Automatic Zooming (SDAZ). SDAZ has been proposed for desktop displays as means of overcoming the problems associated with navigating large information spaces. With this technique, scrolling and zooming are inter-dependently controlled (*AutoZoom*) via a single user action. With the second technique, scrolling and zooming are independently controlled via a single user action (*ManualZoom*). Both techniques are evaluated in a large scale, 72 participant usability experiment alongside a conventional thumbnail-grid image browser. The techniques are tested for their ability to locate *events*, *singles* and *properties* in terms of efficiency (time taken), effectiveness (accuracy) and satisfaction (subjective workloads measured using the NASA Task Load Index [38]). The experiment is carried out on a desktop computer where the viewport size for all three interfaces is set to 240x340 pixels to simulate the display of an HP h5550 Pocket PC and a mouse is used as a stylus surrogate.

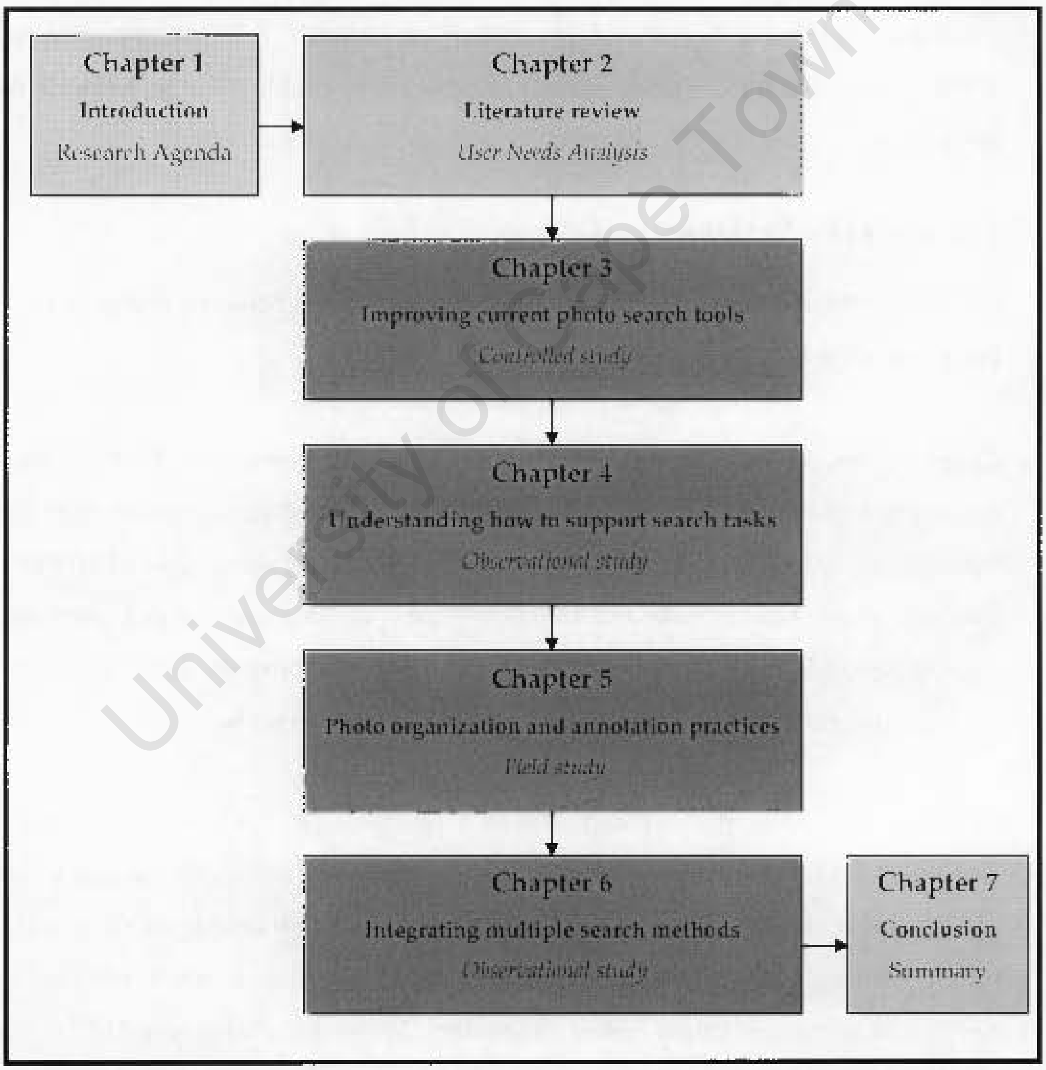


Figure 1.4: Thesis outline. The blue blocks highlight the key research components, describing the key goals and types of evaluation.

Chapter 4 describes how the photo search tool is ported to small display devices. An observational study is conducted to identify problems with the photo search tool so that they can be addressed to provide a better search experience in terms of supporting users in performing the three common searching tasks. The results enable us to obtain a broader view of the design space and allow us to gain some insights into viable solutions. The findings from the study are used to form the requirements for the search tool described in Chapter 6.

Chapter 5 presents a follow-up study that is used to investigate whether people predominately organize photographs by events or by special themed categories. The study also assesses the composition of the annotations that are associated with these picture groupings. The motivation for this study is to clarify some of the observations from the study in Chapter 4. The findings in this study provide an empirical foundation for the query-based search component of the search tool in Chapter 6.

Chapter 6 discusses the design, implementation and evaluation of a photo search tool that integrates multiple methods (*Keyword search*, *Timeline filter*, *Timeline browser*, *Hierarchical folder-based browser* and two *Visual search techniques*) around events. The design rational for this search tool is based on the observations and design implications that are presented in Chapter 4. The results from the study are used to develop a conceptual framework to guide the design of future interfaces based on the amount of information that is known about at event.

Chapter 7 reflects on the major findings and relates them back to our research question. It then goes on to discuss possible directions for further work.

1.9 Definitions

The terminology for photo search research is broad and varied among technical disciplines. For this thesis, we define the following terms:

- *Browsing*: This refers to the process of seeking information when the requirements are not well defined. It requires exposure to the information.

- *Searching*: This refers to the process of seeking information when the search requirements are well defined and can be specified precisely. It can be accomplished by using a query to directly access the data.

- *Small screen device*: This refers to a device that has a low resolution screen, such as a mobile phone or PDA. The term small screen device is used interchangeably with PDA, handheld, mobile device and mobile phone. Table 1.1 shows how current small screen devices have much lower resolution screens in comparison to laptop computers or standard desktop displays. The screen resolutions are compared by reading across a device row to a device column. For example, a Motorola KRZR's screen resolution is equivalent to 1% of the resolution provided by the HP LP3065 LCD Monitor.

		Comparative resolution of devices (%)				
		Motorola KRZR	BlackBerry Curve	Apple iPhone	MacBook Pro	HP LCD Monitor
A	Motorola KRZR Resolution: 176x220 pixels	100	50	25	2	1
B	BlackBerry Curve Resolution: 320x240	198	100	50	4	2
C	Apple iPhone Resolution: 480x320	397	200	100	9	4
D	MacBook Pro, 17-inch Resolution: 1680×1050	4556	2297	1148	100	43
E	HP LP3065 LCD Monitor Resolution: 2560x1600	10579	5333	2667	232	100

Table 1.1: Comparative screen resolutions for a selection of common computing devices

- *Visual photo search technique*: This refers to a search tool where users search for a target by looking through the images in a photo collection. One example is a thumbnail-grid browser.

1.10 Delimitations

The research presented in this thesis is delimited by the photo collection type (i.e. personal photography) and the application domain (i.e. small screen device). The focus is on supporting active searching as opposed to passive searching; where users have to actively acquire information (by searching and browsing) as opposed to passively absorbing information (by monitoring or simply by being aware).

1.11 Summary

This chapter laid the foundations for the thesis. It introduced the research problem and the research questions. The research was then justified, the methodology was discussed, the contributions were described, the structure of the thesis was outlined, the definitions were presented and the delimitations were highlighted.

Chapter 2

Literature review

2.1 *Human Computer Interaction*

The field of Human-Computer Interaction (HCI) is a multi-disciplinary field that draws from computer science, physiology, anthropology, sociology, industrial design, ergonomics and linguistics. Its goal is to facilitate the design, evaluation and implementation of iterative computing systems that satisfy the needs of users [47]. According to Boardman [12],

"A key aim of HCI research is to provide a knowledge base to guide the design of interactive systems. The output of HCI research includes two types of knowledge: substantive, and methodological. Substantive knowledge documents the results of research and may include experimental accounts, the designs of interactive systems and techniques, and models and theories of user behavior. Methodological contributions offer guidance for future research and design in terms of heuristics, tools and methods."

There are two main approaches for identifying user needs and developing design guidelines. The first approach involves modeling user behavior when interacting with a system. These cognitive models can also be used to predict user behavior. One notable example is Card et al's [15] GOMS (Goals, Operators, Methods and Selection rules) model. During tasks analysis, GOMS can be used to determine rules for selecting methods and operations that a user is likely to perform to achieve a goal. Carroll and Campbell [18] have criticized this approach saying that it is too low level, too limited in scope, used too late to influence design and too difficult to apply. Some of these problems stem from the fact that GOMS attempts to predict user behavior by assuming that it remains constant while using a computer system. It is not so straight forward to predict user behavior as it is often shaped by pre-existing mental models [39]. These mental models are not static but are continuously developed during the course of action [92]. Many researchers have proposed extensions to GOMS to address these problems. John and Kieras [54][55] discuss some of these methods and provide guidance to practitioners about which GOMS variant to use for different design situations.

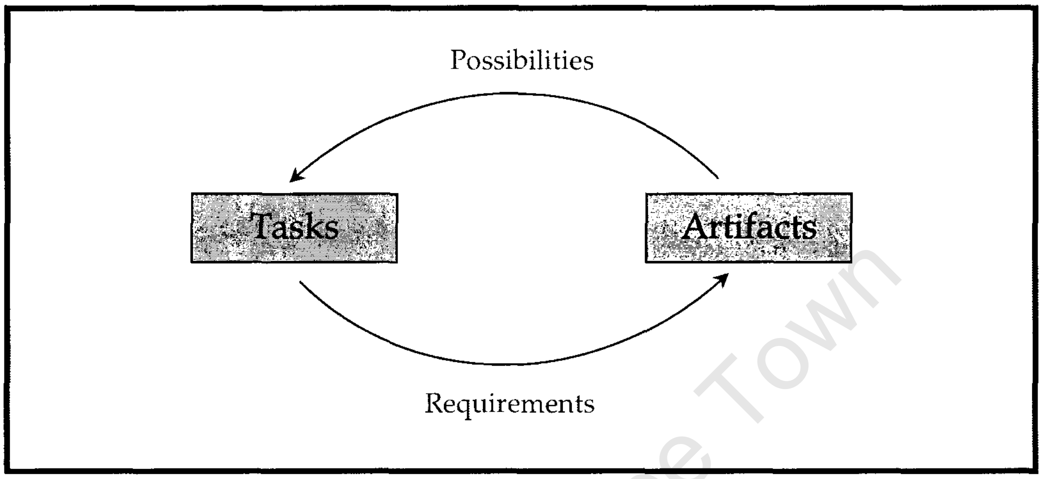


Figure 2.1: The Task-Artifact cycle.

The second approach develops an understanding of user behavior by considering the tools (or artifacts) that are used in an activity and the tasks that users engage in. The task-artifact cycle [17] in Figure 2.1 illustrates this process. The tasks that users engage in define the requirements for an artifact. In turn, this artifact opens up new possibilities for human activities, offering alternative ways to perform familiar activities and enabling new activities as well. These new activities evolve into requirements for the redesign of an artifact. Essentially, the task-artifact cycle is a continuous process of mutually dependent development between a task and an artifact, aimed at reaching an optimal state.

Donald Norman [93][94] emphasizes the need to center the design process around users as they are the ones that use artifacts to accomplish a task. This design philosophy is known as User-Centered Design. Users are involved throughout the design process. An iterative design process is used to obtain user feedback throughout the design and development of an artifact (or system). There are four essential stages: requirements gathering, design, prototyping and evaluation.

At the beginning of the design cycle, the goal is to understand user needs and expectations. A variety of methods such as field studies, focus groups, dairy studies, interviews, or questionnaires can be used to obtain user needs. Based on these initial requirements, low fidelity prototypes are then developed and evaluated. User evaluations and feedback are crucial at this early stage as they allow users to get a better understanding of the system. User needs or requirements that are missed during the initial user needs analysis will often emerge through this process [1]. As the design cycle progresses, higher fidelity prototypes are developed and can be evaluated with more measurable usability criteria such as efficiency, effectiveness and satisfaction. There are generally two approaches to usability evaluation; inspection methods and usability testing. Inspection methods assess the artifact (or system) with out involving users directly [88]. Some examples are heuristic evaluation, cognitive walkthrough and GOMS [31]. Usability testing on the other hand is centered on users and assesses how real users interact with software [88]. The most common techniques are experimental methods, observational methods and query methods [31]. A number of researchers have compared these approaches to evaluation [30][59]. Critical assessments of usability practice can also be found in these papers [14][72]. In addition, Norman [91] and Shneiderman [113] suggest a series of principles that can be used to guide design. Nielsen [88] has developed a practical set of heuristics for usability testing. Hix and Hartson [48] have outlined a practical guide for applying a user-centered design methodology.

A number of studies have reviewed Mobile HCI research methods [10][64]. A key challenge has been how to evaluate mobile systems given that the context is constantly changing [63]. Some researchers have proposed novel solutions for imitating or recreating the context in laboratory settings. For example, by evaluating a mobile system while being simultaneously engaged in a cognitive activity [97].

2.2 *Human factors in photo searching*

Numerous studies [52][61][67][74][123] have discussed how the mobile phone has changed the nature of photography, impacting end-user experiences, expectations and needs. The increased mobility has lead to people taking pictures in situation where they would not have had a digital or film-based camera. The ability to share photographs via MMS or email has enhanced the communicative nature of photographs. Photography has become a social event that is centered on gatherings and having fun [62]. The connection to other people and the capability to entertain them are ends in themselves, where as the utility of the image is often of secondary importance [68]. This interaction is best understood using the classical formulations of sociability, as Simmel [116] explains, "Sociability is interaction that has no external purpose: it is its own goal...Being autotelic, it is very close to games, play and art...The essence of sociability is in the entertainment: the joy an individual gets is fundamentally based on other people also enjoying the interaction. As such, sociability is first and foremost about leisurely conversation, telling jokes and anecdotes."

	Social				Individual	
Affective	**Mutual Experience** Images used to enrich a shared, co-present experience (either in the moment or later as a memento)	103 35%	**Absent Friends or Family** Images used to communicate with absent friends or family (either in the moment or later)	63 21%	**Personal Reflection.** Images used for personal reflection or reminiscing.	120 41%
Functional	**Mutual Task** Images shared with people co-present in support of a task (either in the moment or after the event)	11 4%	**Remote Task.** Images used to help accomplish a task by sharing with remote family, friends or colleagues (either in the moment or later)	28 8%	**Personal Tasks.** Images used to support some future task not involving sharing.	29 10%

Table 2.1: A taxonomy of reasons for image capture [61]

Davis and Van House [123] delineate the social uses of personal mobile photography into five major categories: creating and maintaining social relations, constructing personal and group memory, self-presentation, self-expression, functional and social documentary. Kindberg and et al. [61] present an alternative taxonomy that delineates the reasons for capturing images on a camera phone into two dimensions (see Table 2.1). The first dimension delineates images according to whether they were captured for affective (sentimental or emotional) verses functional (practical or pragmatic) reasons. The second dimension delineates images according to social (sharing with others) verses individual (personal use) intentions. The social category is sub-divided in two, depending on whether images are shared with people that are physically co-present at the time of capture. In total, their taxonomy has six categories, three affective categories: Personal Reflection (41%), Mutual Experience (35%), Absent Friends or Family (21%), and three functional categories: Personal Task (10%), Remote Task (8%) and Mutual Task (4%).

These categories show that although camera phones enable new activities, they do not supplant the traditional activities around photography, such as reminiscing and storytelling. Both these activities are relatively common. In fact, the single biggest category is personal reflection (or reminiscing). When assessing the nature of the images across categories, Kindberg et al. [61] found that people often capture images with their camera phones in places where they often have conventional cameras with them, such as at home, at weddings, or on planned trips. Although, just as many pictures were captured in places where people typically do not have cameras, such as at work or school, in social venues such as pubs and restaurants, or when "out and about," they found that these images did not have a lasting value as they often depict unconventional subjects, which are captured spontaneously for reasons such as amusement, experimentation or curiosity. The more conventional pictures are likely to make up the bulk of the photo collection as these are ones users attach more value to and want to store long term on the mobile device [61]. The new activities introduced by factors such as increased mobility do not have a major impact on this research, as they are more spontaneous in nature, requiring pictures that are captured in the moment. The challenge is in supporting the more traditional activities such as reminiscing which require users to locate older photographs [34], taking into consideration the limitations of current mobile devices such as low resolution screens, varied input mechanisms and limited resources in terms of memory and processing power. Locating older images is not so straight forward as the familiarity of the pictures tends to decay over time. Users are less likely to recall the precise date, the location or even the identities of people in photographs [106].

In a six month study on how people manage their digital photographs Rodden and Wood [105] found that when searching through old personal photographs people are most likely to look for *events* (a set of photographs relating to a particular well defined event); *singles* (an individual remembered photograph), or *properties* (a set of photographs with a common theme). These search tasks are performed irrespective of the reason for searching [105]. For example, to show pictures of previous trip, you need to be able to locate the event; to print a picture, you need to be able to locate a single; to create a collage of a particular person, you need to find a set of pictures of the person (with common property being the person).

As part of the same study, Rodden and Wood [105] wanted to find out the type of features that are commonly used in a digital photo organizer. They developed a system with many features such as audio and text annotation for playback, and content-based image searching. However, users took little advantage of them, emphasizing the utility of two core facilities found in many commercial photo browsers: chronological arrangement and browsable thumbnails. There are three possible reasons for these user preferences: chronological information access is natural for users as shown in the context of email [133] and personal information spaces [70]; users shy away from the computationally expensive content-based image searches, choosing to exploit the human visual system to rapidly scan and process a grid of thumbnails; and, finally, these schemes do not require user effort, like manual annotation, in organizing or pre-processing images.

2.3 Information Access

Broadly speaking, there are two types of information seeking strategies; searching and browsing. When the search requirements are well defined and can be specified precisely, querying provides an efficient way to directly access the data. When the search requirements cannot be expressed precisely, browsing provides an alternative search strategy where users are spared from the difficultly of having to specify a query and can let their search requirements evolve as they browse through the collection. Of course, there is a cost incurred in the amount of time and effort required to browse through the collection until the requirements are met.

Searching and browsing are not mutually exclusive. In fact, the two are intricately linked. Rodden [106] points out that 'there are times when the user might browse through a document to satisfy a requirement, as an alternative to issuing a query. For example, when search requirements are difficult to express'. Likewise, when browsing, it is possible for a search to become more directed as users learn and discover new information. As Cutting and his colleagues [24] explain,

"Access to a document in fact covers an entire spectrum: at one end is a narrowly specified search for a particular document, giving something as specific as a title; at the other end is a browsing session with no well defined goal, satisfying a need to learn more about the document collection. It is common for a session to move across the spectrum, from browsing to search; the user starts with a partially defined goal which is defined as he founds out more about the document collection"

In fact, a study by Bates [7] shows that retrieval strategies are constantly evolving. As the requirements change, users will naturally migrate from one strategy to the other. A plausible intermediary step might be one that combines both search and browsing such as rough searching (which is similar to how people work with search engines like Google). Even when the search requirements cannot be expressed precisely, a query can still be used to reduce the search space to a sensible size. This makes the task of browsing easier as users do not have to sift through the whole photo collection to locate a target.

Chiaramella [19] points out that although a search can provide rapid access to content, it requires a great deal of expertise from the user as they have to master the query language and indexing language. A poor understanding of either is likely to lead to poor results (in terms of precision and recall). Another problem is that users can only view a single set of results at a time. This makes it difficult for the user to infer how previous searches are related to the current one. To construct new queries based on previous results, users are required to memorize information about previous queries and results. This increases the cognitive load on the user. Many of these problems can be overcome by using a browsing-based information retrieval system. These systems require less expertise from the user. They also allow users to interactively access any document in the collection, using predefined access links. The transition from one node to another is based on a proper pre-existing link between the two nodes. However, extensive browsing does have its limitations. It can be time consuming to search large collections, especially when a trial and error search strategy is used. In large collections, another problem can be orientation. The system must ensure that some form of context is maintained. Chiaramella concludes that when considering searching and browsing in terms of cognitive load, bounded or free access to documents and orientation it is evident that they are quite complementary.

2.3.1 Supporting photo search via query-based solutions

The most widely used information seeking tools are web-based search engines such as Google[1], Yahoo[2] or MSN[3] [137]. These systems are able to automatically index documents by extracting keywords from each document. Each unique word is then linked to each document containing it and is given a relevance rank according to a number of factors such as the frequency with it appears in each document and in the collection as a whole. User search terms are then matched against the index, returning the most relevant documents. However, this strategy is less effective for images as 'they do not contain units, like words, that are both meaningful and easy to extract' [106].

Image retrieval systems such as Google Images or Flickr[4] rely on manually annotated data, supplementary text, URLs or filenames to index photographs. However, the search results are not as impressive as those produced by traditional text-based information retrieval systems [53]. A reason for this is that most images on the web do not have structured metadata to describe their content. Any metadata that exists is often inaccurate or incomplete. With personal photography the

[1] Google: http://www.google.com

[2] Yahoo: http://www.yahoo.com

[3] MSN: http://www.msn.com

[4] Flickr: http://www.flickr.com/

contextual metadata is even more sparse [105][106]. Studies have shown that people do not annotate photographs because the benefits of doing so are not always clear [106][111]. Often the annotation task is separate from the tasks that demonstrate the benefit of the metadata [111]. Researchers [38][40][99] often have to rely on lower level metadata, such as EXIF data which encodes the date, time, camera settings, a thumbnail and the location information if it supported by the camera.

Given the rarity of manual annotation, the bulk of research has focused on more automated methods [62]. Instead of relying on textual metadata, content-based image retrieval systems (CBIR) are able to automatically extract and index image content. Similarity comparisons are conducted by comparing features from each image. The most common features are colors, textures and shapes. These belong to the primitive level [106]. Higher level features are closer to user requirements. These belong to the logical and abstract levels. Logical features have to do with the identity of the objects shown on an image, while abstract attributes denote the significance of scenes depicted [106]. Image similarity is usually based on primitive level features such as color. This method is very effective for returning visually similar results such as sunsets. However, people expect similar images to be grouped based on semantic properties, rather than primitive level ones [105]. Unfortunately, content-based retrieval systems are unable to identify and interpret high level features [106][107]. This disparity that exists between the low level features that can be parsed and the semantically meaning descriptions that users want is often called the semantic gap. Although many researchers have attempted to bridge the semantic gap, 'the meaningful identification and interpretation of images remains an AI-complete problem, to which there is no solution in the foreseeable future' [105][106].

Davis and his colleagues [27][131] have tried to overcome this problem by incorporating user feedback in a semi-automated annotation process (see Figure 2.2). Based on the contextual metadata from a cameraphone (e.g. time, location and user name) and the similarity of previously inserted metadata on the server, the system suggests appropriate metadata at the time of capture. Users then have three options: they can accept suggested metadata with a single click or select more appropriate metadata from a list of recommended options or even input new metadata. By incorporating users in the annotation process the goal is to attach more semantically *meaningful* metadata to images. In a study involving 40 participants over a 4 month period they found that users would only input annotations for a few favored images [131]. They also noted that people would have liked to attach labels to groups of related images. This is incredibly important given that this is an activity that users naturally engage in [62]. People tend to think of their photographs in terms of events [62][105]. Unfortunately, this was not emphasized in the system design. Furthermore, the system is reliant on relatively consistent annotation practices between multiple users. This is problematic because annotations tend to be personal. As Rodden [105][106] eloquently describes,

"The annotations will always reflect the subjective judgment and prior knowledge of the annotator, which may not be shared by those searching the collection. In particular, interpretation of the mood or feeling evoked by the image is likely to be highly subjective. Different people notice different things about the content of an image, and have different opinions about what the most important details are…Also it is unlikely that two people would spontaneously use the same term to describe a given concept…In fact, it is likely that annotations made by the same person will be inconsistent between images."

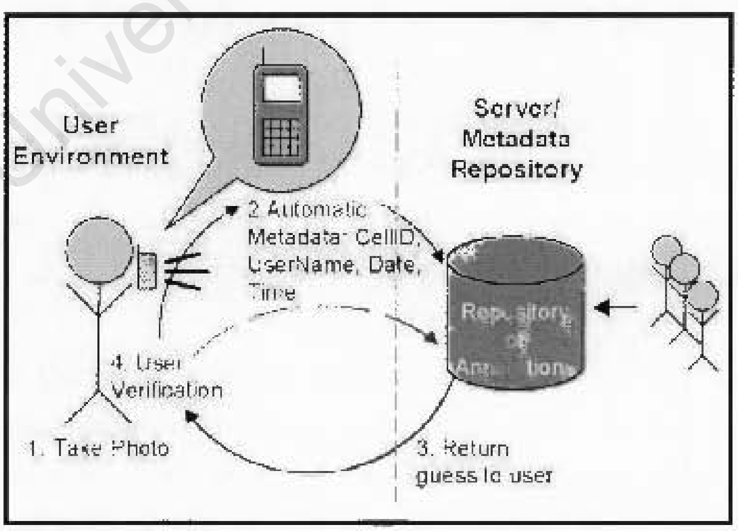


Figure 2.2: An overview of the Mobile Media Metadata (MMM) system which incorporates user feedback in a semi-automated annotation process. [131]

There are other annotation approaches. One particularly innovative approach proposed by Kang and Shneiderman [58] allows users to simply drag and drop annotations on to one or more photographs. For example, by dragging a number of contacts on to a photograph to indicate who is in the picture. However, no user studies have been conducted to see whether this approach motivates users to annotate the bulk of their photographs.

Several other attempts using audio annotation [36][106] have been made to solve this problem but ultimately their uptake has been poor, due to the usability issues associated with complex or novel user interfaces and also the fact that most users simply do not wish to change the way they work.

2.3.2 Supporting photo search via clustering and visual solutions

Many commercial photo management tools provide mechanisms for navigating through a photo collection (e.g. a flat scrollable thumbnail-grid browser. Search tools that allow users to browse through the photo collection are versatile in supporting a range of user needs. As Korn explains,

"Browsing is useful and important in information retrieval. While queries are powerful methods for specific searches, they are not useful to the user when the information sought is vague and ill-formed or when the user can only describe it in the right context. Often the user, being unaware of the database structure, has no idea whether the search term has an appropriate scope. Browsing on the other hand is more exploratory and interactive and allows users to apply recognition skills to match target concepts rather than recalling specific items. ***Depending on the task at hand, browsing can take on a range of different functions, from specific to purposeful to serendipitous.****"* [66]

Two photo search tool design approaches that have been used to support users when browsing through their photographs are clustering and visualization. Clustering techniques group images into various categories by imposing a top level structure on the data. Visualization techniques enable users to visually search through images in the photo collection. Visual photo search techniques display images in various configurations and provide controls for navigating through the photo collection.

Clustering

The Timeline browser [44] is an example of a system that automatically clusters photographs into a hierarchical time-based structure (see Figure 2.3). It has three views: a year view, a month view and a thumbnail view. The photographs are automatically placed into system generated clusters (or albums). Tapping on an album enables you to drill down into a finer time range, until the thumbnail view is displayed.

Figure 2.3: Timeline browser showing the year view (left), four month view (middle) and the thumbnail view (right) [44].

The Timeline browser was compared against a standard thumbnail browser [44]. Participants were asked to locate individual photographs and also find some pictures for a collage. They found no significant differences between the two systems in terms of the time taken to complete the tasks. There were also no significant differences between any of the subjective measures. They did, however, find that users were able to complete tasks more successfully with the Timeline browser. One plausible explanation for the relative lack of performance and poor subjective ratings is that the overall design of the Timeline browser over looks the fact that events are most natural way of thinking about photographs [62][105].

Photographs naturally cluster around events. However, these natural event clusters can get broken up into multiple time-based clusters. When looking at a time-based cluster it is not immediately obvious which events it contains. In addition, the added difficulty of having to navigate in and out of artificially created clusters can increase the cognitive load on the user.

Other approaches automatically cluster photographs into events based on visual patterns [99]. This is also problematic because the definition of an event is very subjective. It is unlikely that these event clusters will always match what people perceive to be an event. Any mismatch is likely to complicate navigation.

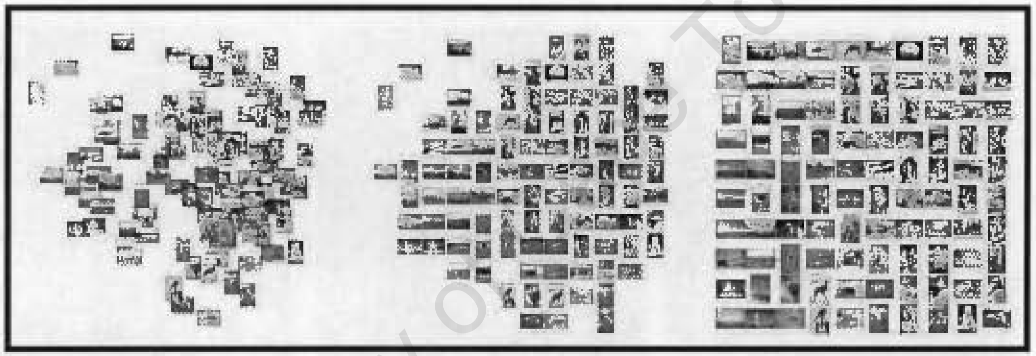


Figure 2.4: Organizing images according to visual similarity with overlapping (left), no over lapping (middle), and maximizing thumbnail size (right) [107].

Rodden and her colleagues [107] have looked in to clustering pictures according to visual similarity (see Figure 2.4). They concede that to cluster images according to their semantic similarities requires meaningful feature-based extraction and identification which has yet to be solved convincingly by AI research. Furthermore, this technique will not scale as a photo collection grows in size as it relies on showing all the pictures at once. This is more significant on handheld devices where only a few photographs can be shown at a time.

Photographs have also been clustered by location [6][86]. Kirk and his colleagues [62] question the need for such an approach. In their study they observed that people organize photographs into folders according to events. Each folder is often labeled with a meaningful description and date, both of which users are heavily reliant on this information when searching their photo collection. 'The fact that users do this naturally casts doubt on the need for photo grouping algorithms which are being developed to cluster pictures using GPS data, as grouping by time and event also tends to group by location as a by-product' [62]. Furthermore, the location of the photographer can differ by several kilometers from the location of the photographed subject [26]. There are also technological problems. The cell information from a cameraphone is not precise. Although GPS is more precise, it is unreliable indoors. GPS enabled cameras are also not wide-spread. That said, location information can still be helpful for distinguishing events.

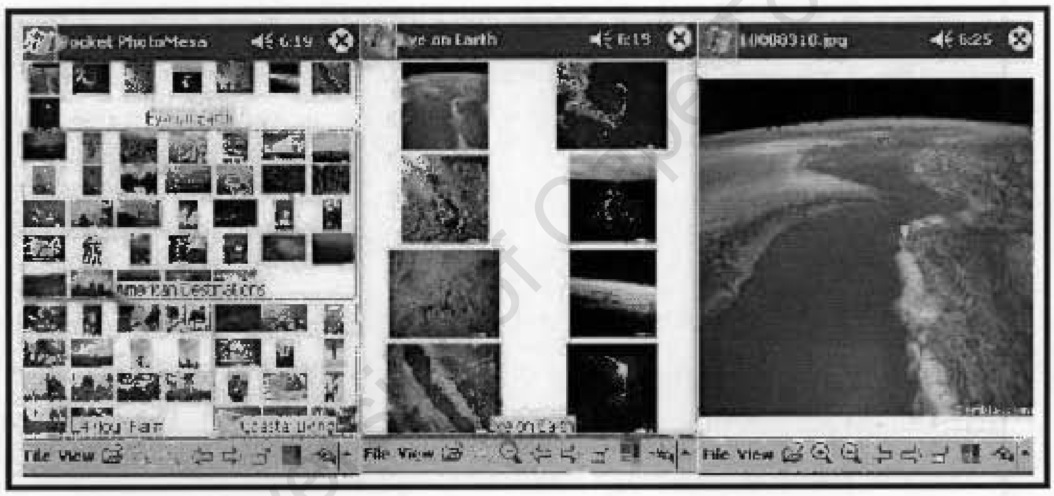


Figure 2.5: Three views of Pocket PhotoMesa browser at different zoom levels [65].

Bederson and Khella [65] provide a visually appealing solution that uses quantum treemaps to display categories of images in the most screen efficient way (see Figure 2.5). Quantum treemaps are a variation of existing treemaps algorithms that are designed for laying out images or other objects of indivisible (quantum) size. The problem with existing treemap algorithms is that they all return regions of arbitrary aspect ratios. Quantum treemaps guarantee that the regions that show groups of photos have dimensions that are integer multiples of the dimensions of the photos. This minimizes the amount of space that is wasted in each region. By default, the Pocket PhotoMesa browser clusters images according to directories. It can also be configured to cluster images according to other user defined metadata. The thumbnail size is adjusted accordingly so that all the photographs can be shown concurrently. When a category is selected, the view is smoothly animated in to one of several predefined zoom levels. When a thumbnail is selected the image is smoothly animated in to fill the screen. The user can zoom in further to view portions of the image or pan around the image.

In a user study the Pocket PhotoMesa browser was not able to locate images any faster than the baseline browser (Pocket ACDSee[1], a commercial thumbnail browser). It did however score higher in the subjective ratings, mainly due to the smooth animation and the treemap layout. The authors concede that this solution is perhaps better suited for navigating and exploring rather than locating. The fact that users had difficulties recognizing thumbnails when 75 images where shown concurrently suggests that this technique is unsuitable for image collections that are any larger than this.

Visualization

Given the scalability problem of displaying images concurrently on a handheld device, one solution is to trade off space for time, by showing portions of the photo collection over a period of time. Special controls allow users to search through the contents, by enabling them to navigate through the photo collection. An example of such a technique is Rapid Serial Visual Presentation (RSVP) [119], where the most common analogy is that of riffling through a book where momentarily seeing each page allows you to assess the content. Other presentation modes allow images to be displayed at several fixed locations on the screen [118].

Apple's iPod Photo [4] provides a photo search solution for rapidly scrolling through photographs. It overcomes the screen size limitations by trading space for time. Once an album is selected, the photographs are displayed in a 5x5 grid of thumbnails (see Figure 2.6). A thumbnail can be enlarged to fill the screen. The jog-

[1]http://www.soft32.com/download_15051.html

wheel is used to rapidly scroll through the thumbnails. It can also be used the rapidly scroll through full sized images.

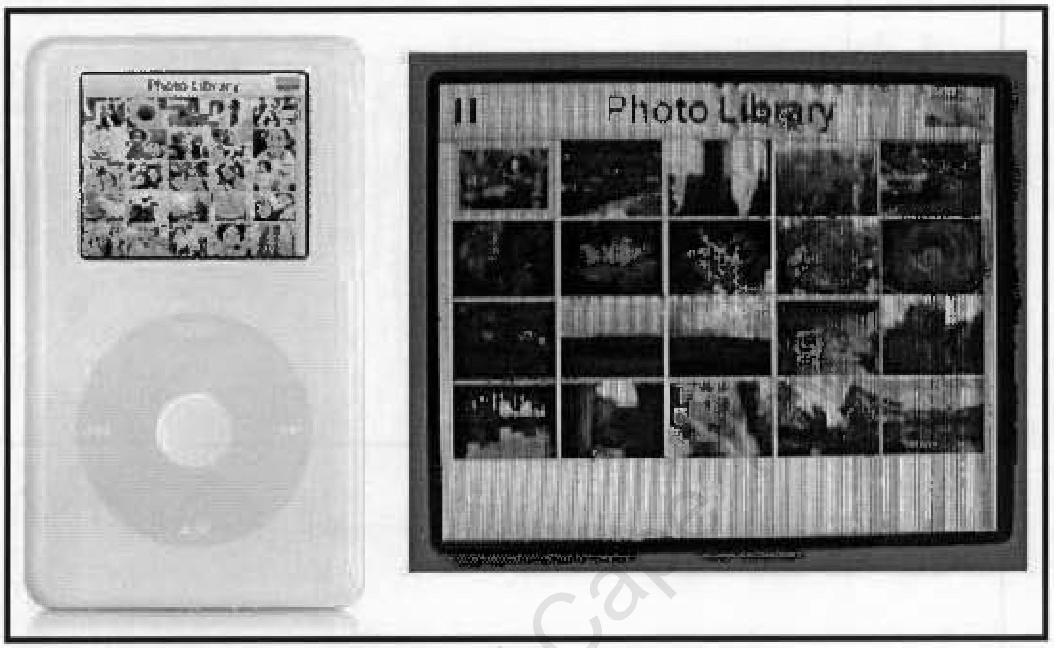


Figure 2.6: iPod Photo (left), close up of the screen (right).

Marsden and Jones [56] express disappointment with the lack of smooth transitions between the two presentation modes (the thumbnail view and the full-sized image view) to maintain the context. Another concern is that the thumbnails are too small to let you let you distinguish variations of the same shot. For example, a series of pictures that are taken with a camera's rapid shot feature. The lack of intermediary thumbnail sizes makes it is difficult to compare neighboring photographs.

Derthick [29] ran a number of studies to investigate the effectiveness of parallel presentation (showing a grid of thumbnails) versus serial presentation (showing a single image) using two control modalities, scrolling and RSVP. He developed four interfaces: parallel presentation using scrolling (PS), parallel presentation using RSVP, serial presentation using scrolling (SS) and serial presentation using RSVP. He found that the PS technique was significantly faster than SS technique at locating a target image from a collection of 200. However, there were no other significant differences. Interestingly, his participants preferred scrolling interfaces to the RSVP ones. They also preferred serial presentations over parallel presentations.

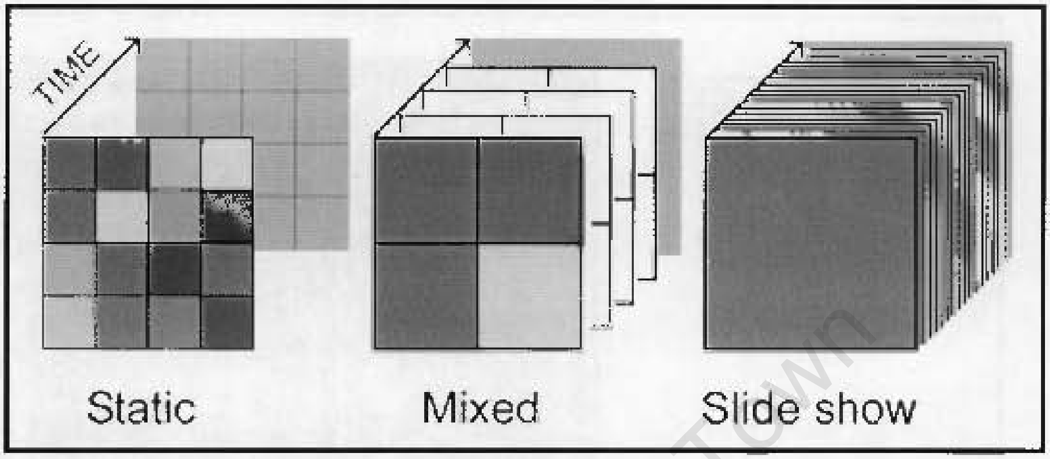


Figure 2.7: Three schematic RSVP presentation modes [118].

Spence and his colleagues [118] conducted a much more rigorous experimental study to identify the best presentation mode for displaying images (see Figure 2.7). They created three modes: *static* (showing 16 images concurrently), *mixed* (showing 4 images concurrently) and *slideshow* (showing a single image at a time). The presentation time was dependent on the number of images shown concurrently. For example, if the duration for a single image was 100ms, the *static* mode was displayed for 1.6 seconds, the *mixed* for 400ms and the *slideshow* for 100ms. For each presentation mode, users were required to locate a single target image from a collection of 64, shown at four different presentation rates, using three different screen sizes (simulating a desktop computer screen, a PDA screen and a mobile phone screen). In their study, users found it difficult to identify images that were displayed for less than 100ms. They found the *slideshow* mode particularly difficult when the presentation time was less than 100ms. Overall, the *mixed* mode was the least error prone. It was also the most preferred mode.

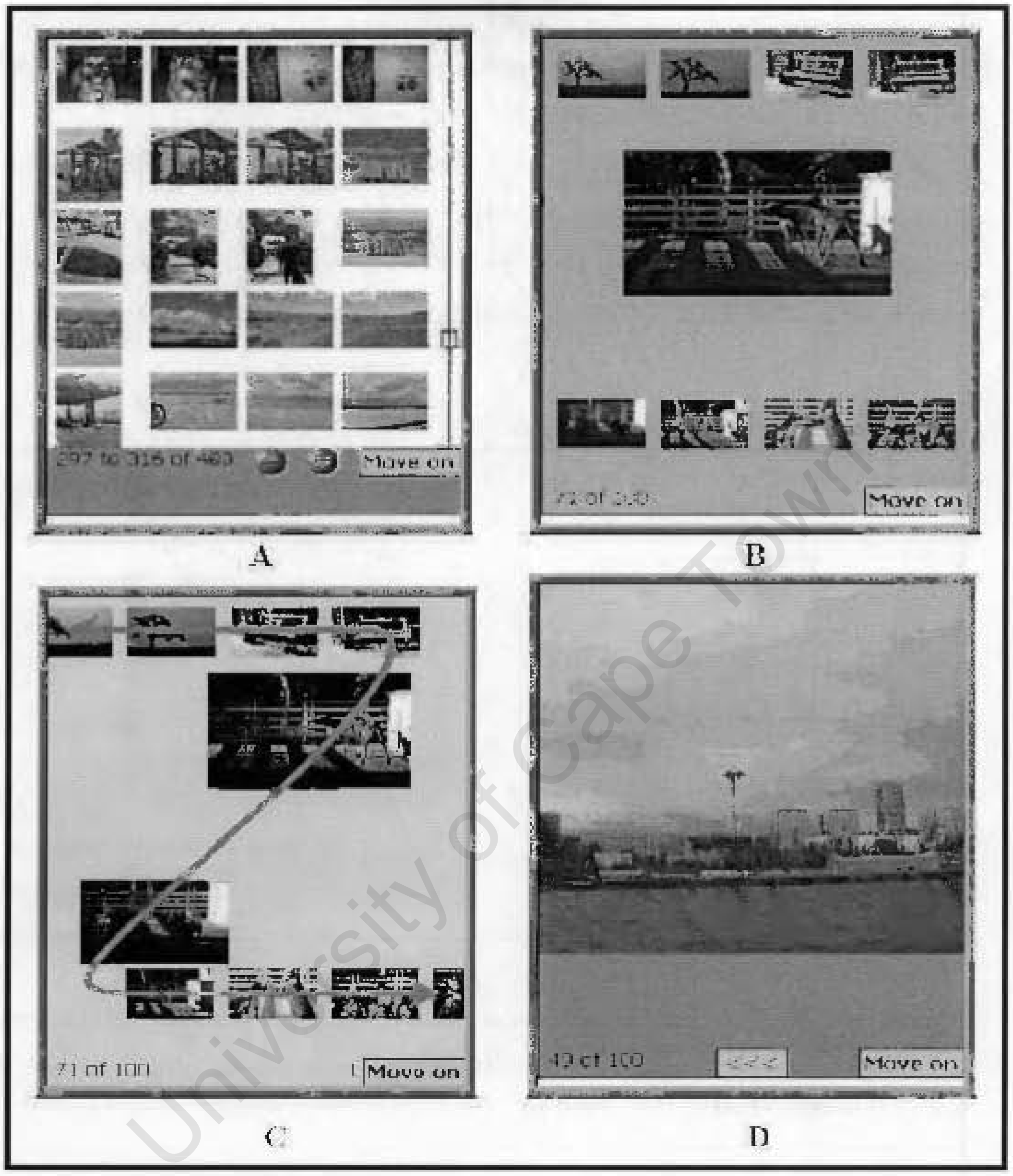


Figure 2.8: (A) *Thumbnail* layout, (B) *Flip-zooming* layout, (C) photos animate along the "Z" shaped motion path. The drawn-in path in the figure is not visible on the actual interface, and (D) the *RSVP* layout [128].

Wang and his colleagues [128] recently conducted a study to find the optimal combination between presentation modes and physical control modalities. They developed a total of nine different visual photo search tools consisting of three presentation modes and three control modalities (see Figure 2.8). The three presentation modes were a traditional *Thumbnail* layout, a *FlipZoom* layout and an *RSVP* layout showing a single image at a time. The three physical control modalities were a *jog-dial*, a *squeeze sensor* and a *click-based on-screen control.*

They developed a model to categorize the search tools according to *Cognition* and *Manipulation* (see Figure 2.9). *Cognition* refers to how well the user is able to comprehend what is on the display. *Manipulation* refers to how effectively the user is able to manipulate the control modality.

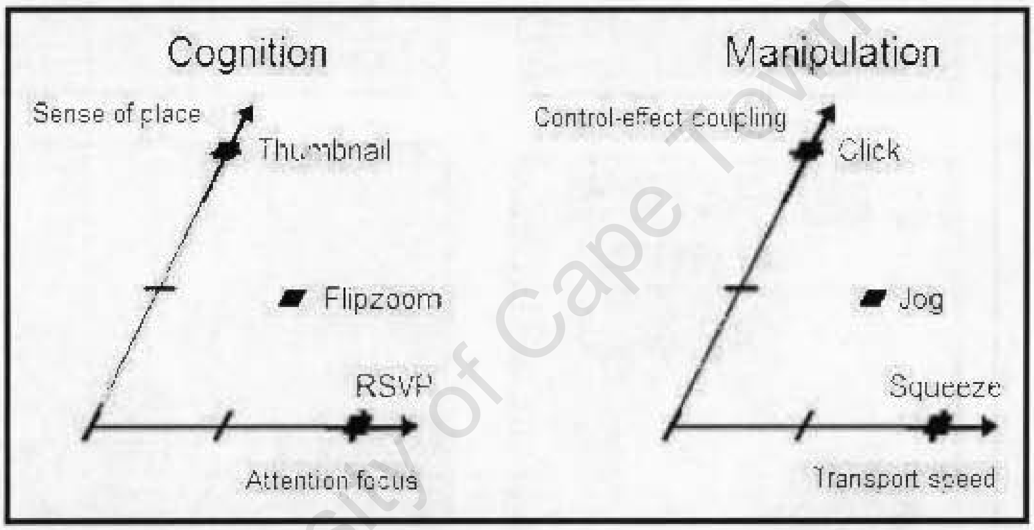


Figure 2.9: ***Cognition–Manipulation*** **framework.** ***Cognition*** **refers to how well the user is able to comprehend what is on the display.** ***Manipulation*** **refers to how effectively the user is able to manipulate the control modality [128].**

In terms of *Cognition,* they focus on two factors that are important for image searching; the user's sense of place (context) and the degree of attention focus an interface elicits. The *Thumbnail* layout provides significant sense of place; however the viewer is required to split his attention across the entire screen. The *RSVP* layout allows the user to examine a single photograph almost immediately, allowing a high attention focus at the expense of context (or sense of place). The *FlipZoom* layout is a compromise between the two.

In terms of *Manipulation* they focus on two factors; the transport speed and the control-effect coupling. The *squeeze* modality allows the user to scroll through the images at various user controlled speeds. It also has the loosest coupling, since it provides the most indirect control, mapping the pressure exerted on the sensor to the scroll speed. With the *click* modality users have to explicitly click a "next" or "previous" button to scroll through a photo collection. The transport speed is also limited by the users tapping ability. The *click* modality has a high control effect coupling as each click advances photos by one unit. The *jog-dial* modality provides a compromise between the two. Its on/off nature allows it to be mapped to scroll speeds the average user can follow. One limitation is that users cannot accelerate to rapidly scroll through irrelevant pictures or decelerate when approaching more relevant pictures. In terms of control-effect coupling, it is placed between the two extremities. It is less coupled than the *click* modality as you can hold the *jog-dial* in the up or down position, allowing scrolling to continue without any further user input. It is more coupled than the *squeeze* modality because of its on/off nature.

For both *Cognition* and *Manipulation*, a high value on either axis is desirable. However, they are mutually conflicting. For example, a high attention focus requires sacrifices in sense of place. Similarly, high transport speeds can lead to a feeling of less control. The interaction between *Cognition* and *Manipulation* is equally important to consider and understand. For example, the *Thumbnail* modality is best coupled with a lower transport speed, due to its low attention focus and resulting absorption delays, whereas the *RSVP* modality is likely to require a higher transport speed.

To investigate all these issues, Wang and his colleagues [128] conducted a multi-factor experiment. They found no significant differences between the three presentation modes in terms of the time taken to locate photographs and the success rate in completing tasks. The *RSVP* layout was the most preferred presentation mode. Participants felt that it had the most appropriate photo size. Overall, the *Jog-dial* modality performed the best. This result was somewhat surprising for the authors as they expected the *Squeeze* modality to out-perform the rest. The *Squeeze* modality was actually perceived to be the least reliable due to this dynamic control.

The authors felt that this was due to the pressure sensor being too sensitive. Consequently users would frequently overshoot target photographs. In terms of their model, they suspect that the control-effect coupling for the *Squeeze* modality is currently not tight enough. However, the authors concede that it would be difficult to find a single operating parameter to suit all users. They recommend the need to personalize the slope of the pressure/speed function automatically.

All these studies point to the need for dynamic control of the presentation mode and the control modality. Spence and his colleagues [118] found that the *Mixed* presentation mode (showing a few large images) was the least error prone and the most preferred mode. This is because it provides the best compromise between context and detail. In the *Slideshow* mode (detail view), it is difficult to compare adjacent images as only a single image is shown at a time. It is not surprising that this mode was the most error prone as it suffers from a lack of context. In the *Fixed* mode (thumbnail view), photographs are too small to distinguish features between similar photographs. Ideally users should be able to adjust the level of detail that is shown to suit their informational needs. Similarly, in Derthick's study [29] users preferred scrolling to RSVP. The scrolling interface allowed more dynamic control, enabling users to rapidly navigate towards a target photo and slow down on reaching neighboring images, whereas with the RSVP interface a constant presentation rate was maintained. Providing dynamic control of the presentation rate is not enough. Wang and his colleagues point out the need to automatically customize controls to suit user needs.

2.4 *Designing visual photo search tools for small screen devices*

The discussion on visual photo search tools thus far has focused on presentation modes and control modalities. It is also important to consider the cognitive processes and activities users are engaged in when navigating through a photo collection.

2.4.1 A framework for navigation

Several researchers [8][120][130] have proposed general frameworks for decomposing navigational tasks into their constituent parts. Spence's model [120] is applicable to a whole range of navigation tasks. His framework decomposes navigation into four cognitive activities: searching and browsing; formation of the internal model; interpretation of the internal model and formation of a retrieval strategy (see Figure 2.10). Together, these stages form a single cycle of an iterative navigation process.

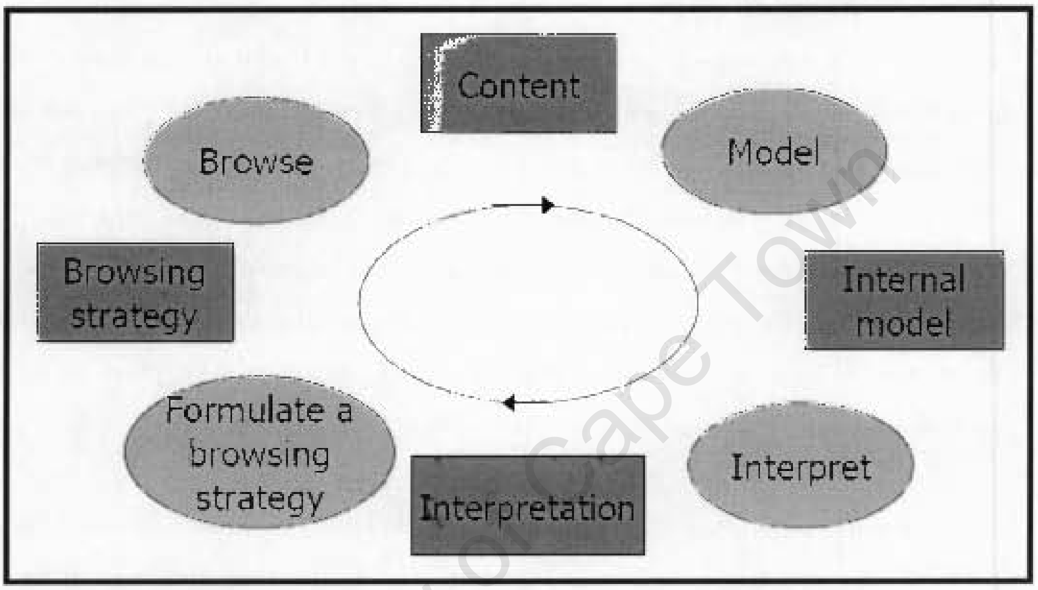


Figure 2.10: Spence's framework for navigation [28]

In the *browsing* stage, a user looks through the content to see what is there. In the *model* stage, content acquired by browsing is used in the formation of an internal model (cognitive map). In the *interpretation* stage, the internal model is used to decide whether the goal had been reached or whether the browsing strategy should be revised. In the *formulate a browsing strategy* stage, the browsing strategy is revised and a new direction for browsing is chosen. This process can be cognitive (driven by the interpretation of a new idea) or perceptual (influenced by what is displayed).

De Bruijn and Spence [28] used this framework to investigate the consequences of the space-time trade-off for information transfer. Their primary focus was on the first two stages of the framework (searching and browsing, and the formation of an internal model) as these stages have the biggest impact on the usability of a navigation-based technique. They state that the limiting factor for techniques that trade-off space for time is the human visual system, as there is a lower limit for presentation time beyond which an item cannot be processed fully. This limit is not fixed, but depends largely on the nature of the task. Designing an effective navigation technique requires careful consideration of this factor.

2.4.2 Design implications for searching

For searching, when the target images are known, the visual system may perform attentional selection, where potential targets are processed more fully than others in the stream [101][115]. This means that when locating a target image the presentation time can be much less than the total time needed to process each image fully [28]. The search time depends on whether the attentional selection takes place based on visual cues (e.g. the appearance is known) or categorical cues (e.g. searching based on a description), as selection based on visual cues can take place at higher presentation rates than selection based on categorical cues [28].

Personal photographs are often captured in bursts [40]. The time difference between photographs in a burst (or cluster) tends to be smaller than the time difference between clusters [40]. Rodden and Wood observed that photographs that taken at the same time are often visually similar [87]. These features can be exploited to allow higher presentation rates by giving the illusion of a moving coherent scene as in cinema [25]. This means that each image does not need to be processed fully as additional information requirements can be obtained from subsequent frames [122]. However, if adjacent images are not similar, maintaining the same rate will result in images not being fully resolved.

If the user is searching for multiple items, the identification of one may interfere with the identification of subsequent targets [13]. This phenomenon is known as the attentional blink [101]. It lasts for several hundred milliseconds after the identification of the first target [101]. The consequence of this is that subsequent targets may not be processed fully. For example, if a target image appears close to a landmark, the identification of the landmark image may interfere with the identification of the target image [28].

2.4.3 Design implications for browsing

When browsing, users start with a partially defined goal and look through the content to see what is there. As they acquire more information by browsing, Spence [28] notes that an internal model is formed and refined through an iterative navigation process. The overall objective is to define the goal and then locate the target. To assist users in the formation of an internal model the user must be able to process and store information on each image. Therefore the effectiveness of a navigation technique is also dependent on how well users can recall items in an information stream. The ability to memorize items is dependent on factors such as the similarity of photographs and the presentation time [28]. As the presentation time decreases the ability to memorize items diminishes dramatically, from very good when the presentation time is set to 5 seconds to very poor when the presentation time is only 100ms [51]. This suggests that not more than 10 images should be shown each second.

However, there are number of studies that suggest that more information about individual items in a stream is being processed than is apparent from user performance in memory recognition tests [28][77][122]. It may be the case that sufficient information is processed to create an internal model of the content, even though individual items are not fully processed. These findings are supported by Tse and his colleagues [122], who found that users are able to determine the gist of videos after being shown key frames at a presentation rate of 16 frames per second using RSVP. However, their recall of individual frames was poor. Furthermore, Marcel [77] reports that when showing information for a brief period, more information is available than what can be reported by the observer.

2.5 *Summary*

Section 2.1 introduces the field of Human-Computer Interaction. Section 2.2 discusses the human factors in photo searching. Section 2.3 discusses the linkages between two methods of information access, searching and browsing. It also conducts a critical analysis of current photo search solutions. Section 2.4 uses framework for navigation to describe some design implications for photo search tools. This section reviews the background literature and discusses the course of research that is followed.

This chapter highlights a number of user needs and the inability of currents systems to support them. When searching for photographs, people are likely to look for *events*, *singles* and *properties* [105]. Depending on how well-defined the search requirements are, they might elect to search or browse. When the search requirements can be specified precisely, searching (or querying) provides an efficient way to directly access the data. When the search requirements cannot be expressed precisely, browsing provides an alternative search strategy where users are spared from the difficultly of having to specify a query and can let their search requirements evolve as they browse through the collection. Searching solutions have been relatively unsuccessful due to the lack of meaningful metadata for each image. Browsing solutions on the other hand do not rely on user annotations. They make use of human visual systems ability to process visual information quickly. However, current browsing solutions are not flexible enough to automatically adjust the presentation mode or the presentation rate to meet user needs and do not make optimal use of the human visual system.

The following chapters address the attributes of searching and browsing that are not well supported for photo search. In Chapter 3, a search tool is developed to address the limitations of current visual photo search tools on small display devices such as the lack of flexibility in configuring the presentation mode and the lack of dynamic control of the presentation rate. Two different techniques are developed to address these issues. One approach automatically configures the presentation mode and presentation rate based on user actions. The other approach provides separate controls for the presentation mode and presentation rate. The search tool is evaluated against a thumbnail-grid browser to see whether the new techniques provide a significant improvement on current photo search solutions. In Chapter 4, an observational study is conducted to see what is wrong with existing visual photo search tools so they can be addressed to provide a better search experience in terms the common photo search tasks, locating *events*, *singles* and *properties*. Some design implications are proposed to further improve the tool. In Chapter 5, a small follow-up study is conducted to clarify some of the observations that are made in Chapter 4. More specifically, to investigate whether people predominantly organize photographs by events and to obtain an accurate reflection of the type of metadata that is associated with events. One major finding is that people engage in a more light weight form of annotation that is centered on events. This particularly important because it provides an alternative solution for supporting query-based searching that does not rely on having metadata for each image. Query-based photo search tools can be designed to locate events rather than individual images. In Chapter 6, a search tool is developed based on the user observations and design implications presented in Chapter 4 which emphasize two core requirements: firstly, the need for the search tool to locate events rapidly and secondly, the need to provide multiple methods to locate events. The search tool integrates multiple search techniques around events. These include a query-based technique and the visual photo search techniques presented in Chapters 3 and 4. Some design guidelines are proposed for future systems based on a user evaluation of this tool.

Chapter 3

Improving current photo search tools

3.1 Introduction

The aim of this chapter is develop a visual photo search tool for small display devices. The requirements for this tool are based on the limitations of current visual photo search tools which were identified in the previous chapter. To support users needs more adequately these tools must:

- *Allow the presentation mode to be dynamically configurable.* The visual photo search techniques presented in this chapter allow the presentation mode to be automatically adjusted so that the appropriate level of detail can be selected.

- *Provide a control modality that dynamically adjusts the presentation rate.* The visual photo search techniques also allow users to adjust the presentation rate. This is essential for supporting attentional selection (see Chapter 2.4.2) and also assisting users in processing and storing information on each image (see Chapter 2.4.3).

- *Support personalization.* The models and algorithms used are kept as simple as possible to make it easy to customize the search techniques.

- *Make optimal use of the human visual system.* Two different approaches are used to make optimal use of the human visual system. One approach automatically adjusts the visual flow of information to maintain to optimal rate at which the human visual system can process information. It prevents the visual system from being overloaded. The other approach provides manual control of the visual flow.

The design approach that is used is to develop the tool is deliberately incremental rather than revolutionary. Whitaker, Terveen and Nardi [132] note that while revolutionary design is vital for making progress, designers should always try to improve existing work. Only when prior work can no longer be improved does it makes sense to use a revolutionary approach. Boardman [12] notes that "a radical invention does not necessarily entail a strong contribution in terms of HCI knowledge, especially if no evaluation has been performed." Incremental designs allow direct comparisons to be made with existing work. They promote system uptake and also take advantage of user familiarity. Following an incremental approach does not mean further more ambitious changes can not be made. Once the initial changes have been evaluated, subsequent refinements can be made as part of an iterative user-centered design process. The effects of any refinements are scrutinized during a design cycle and are more easily understood because they can be compared to a previous design [132]. This evolutionary design facilitates the creation of knowledge.

3.1.1 Outline

This chapter presents the incremental design, implementation and evaluation of a photo search tool that supports three different visual photo search techniques, two novel techniques and one more traditional technique. Section 3.2 provides a critical analysis of different methods that allow users to visually search through large information spaces. Speed Dependent Automatic Zooming (SDAZ) is identified as a suitable approach, due to its ability to address the limitations highlighted in Section 3.1. Section 3.3 discusses the mechanics of the SDAZ technique. Section 3.4 discusses some algorithmic extensions that are made to SDAZ to make it more suitable for small displays. Section 3.5 discusses how this technique is used to create two novel visual photo search techniques. It also discusses the design and development of a traditional visual photo search interface. Section 3.6 evaluates the new interfaces against the traditional interface in terms of user performance and subjective preferences. Section 3.7 summarizes the major findings and provides an outlook for subsequent chapters.

3.1.2 Contributions

This chapter makes two contributions to this thesis:

- *Design, implementation and evaluation of two visual photo search techniques that allow users to visually search through photographs on small displays.* One technique is an adaptation of a navigation technique called Speed Dependant Automatic Zooming [49]. SDAZ has been proposed for desktop displays as means of overcoming problems associated with navigating large information spaces. With this technique scrolling and zooming are inter-dependently controlled. With the second technique, scrolling and zooming are independently controlled via a single user action. The evaluation shows that the new techniques support faster navigation to target photographs. They also support more accurate identification of photographs. The subjective workload is also lower for the new techniques. The two visual photo search techniques are distinguished from previous research because they take into account the visual searching needs that are outlined in Section 3.1. The techniques provide dynamic control of the presentation mode and the presentation rate. The also make optimal use of the human visual system and support personalization.

- *Algorithmic extensions to SDAZ.* Four main extensions are made to SDAZ. The first extension factors in the proportion of the information space that is visible on the screen. The second extension proposes a method for mapping the scroll speed to the zooming level based on horizontal and vertical translations in the information space. The third extension generalizes the algorithms to support navigation both horizontally and vertically or constrained to one of the two dimensions. The fourth extension addresses scrolling and zooming behavior when the bounds of the information space are reached. Previous research on SDAZ has only provided partial explanations of how to implement the technique, choosing to focus more on the interaction mechanism and its effect on users. This chapter provides a clear implementation guide for developers and practitioners

3.2 Background

'A problem humans experience in their everyday lives is having too many things to put in a limited space: furniture in a house, books on shelves, windows on a computer screen, data to display on a PDA' [78]. The information explosion in the last hundred years has created information spaces that are often too large to be displayed on a single screen. Documents, web pages, spread sheets, and pictures are examples of information spaces that commonly suffer from this problem.

One way of dealing with this presentation problem involves showing a portion of the information space on the screen and providing controls to navigate through the information space. Some common controls are scrolling, panning and zooming. Scrolling and panning are used to bring data in and out of the visible area. Zooming is used to increase or decrease the size of the information space that is visible on the screen.

Although many applications adopt these basic functions, there are number limitations in the navigation they provide. Igarashi and Hinckley [49] note that when browsing a document, users have to shift their focus between the document and the scrollbar. This causes an attentional overhead and also increases the operational time. They also observe that when browsing large documents, small scrollbar movements result in large movements of the document. This can disorientate and confuse users. Cockburn and Savage [21] note that because zooming changes the size of the information space that is visible, more scrolling is required when zoomed in and less when zoomed out to navigate the same distance. In order to predict scroll bar interactions, user must understand the relationship between the scroll distance and the zoom level. Furthermore, when panning is not supported in an application, users are restricted to independent horizontal and vertical controls. This requires a greater number of interface actions as navigation in other directions (such as diagonally) requires at least two scrolling operations. These problems are of even greater concern on small display devices, such as PDAs and mobile phones as the portion of the information space that is visible is much smaller than standard desktop displays. The implication is that more scrolling, panning and zooming is required to navigate information spaces when conventional navigation mechanisms are used.

Two classic solutions to this problem are overview-and-detail and focus-and-context. The overview-detail approach simultaneously displays two separate views, one for context and the other for detail. Chittaro [20] notes that 'although this approach is feasible on desktop computers, it tends to fail on mobile devices. The limited screen space makes it difficult to relate the two views, and it can even be impossible to properly fit them on the same screen.' This approach also forces users to mentally integrate the context and detail views. An operational overhead is also incurred in moving between the two views. Another design approach for the overview-and-detail design pattern on mobile devices is a one-window-drill down, where users are initially shown an overview and can drill down to view detail. This approach maximizes the use of the limited screen real-estate and can be particularly effective in guiding navigation when the overview displays objects that are of interest to the user. One reason why this approach has been used less on mobile devices is due to the scalability issue which relates directly to the size of the overview. When the zoom factor is high, the overview can shrink too much in size, obscuring points of interest and making manipulation difficult for users.

The focus-context approach integrates the overview and detail within a single view. One example of this approach is the fish-eye view which magnifies objects at the point of focus and progressively decreases the size of more distant objects. Implementing an efficient fish-eye view on a mobile device in neither easy nor effective [103]. The granularity of the control decreases as the information space gets larger, making it harder to locate a target object. The gradually increasing magnification makes targets move faster and faster the closer they are to the focal point. This makes if difficult to acquire targets as they moving fastest when the point of focus is directly over the target [41].

Another approach for addressing the limitations of conventional navigation mechanisms retains traditional scroll and zoom functions but provides alternative input mechanisms. For example, some small screen devices have wheel controls or joysticks that can be used for scrolling, to alleviate some of the associated problems, such as the attentional overhead. One attendant problem though is that as the scroll-rate increases, the flow of information becomes too great for users to perceive, resulting in visual blurring and consequent user disorientation.

To counter this visual blurring Igarashi and Hinckley [41] proposed Speed Dependent Automatic Zooming. As the scroll speed increases the information space is zoomed out to reduce the visual flow (or rate of pixel movement). This allows users to scroll more rapidly through an information space without overloading the visual system. The level of detail that is shown is automatically adjusted according to the scroll speed, showing the detail view when scrolling slowly and progressively showing less detail as the scroll speed is increased. The smooth transition between the global overview and the magnified local view assists users in building a mental map of the information space. The simplicity of technique makes it easy to customize to meet user needs.

However, SDAZ was originally proposed on for desktop displays as means of overcoming the problems associated with navigating large information spaces. There is a need to adapt this technique to the small screen. Before doing so, it is necessary to examine the mechanics of SDAZ.

3.3 Speed dependent automatic zooming

Igarashi and Hinckley [49] use the following equations to compute speed and scale:

$$speed = C * dy$$

Equation 1

In Equation 1 *dy* denotes the displacement of pointing device such as a mouse. C is a constant that is used to adjust the scroll speed. The scale is computed using the following equation:

$$scale = \frac{v0}{speed}$$

Equation 2

where $v0$ is a constant that represents the scroll speed when zooming commences (see Figure 3.1a). However, there are two concerns with Equation 2. Firstly, no limits are provided, so when the $speed \to 0$, $scale \to \infty$. This contradicts Figure 3.1a which shows that when the $speed = 0$, $scale = 1$. Secondly, as the scroll speed increases this formula causes a sudden drop in the scale and then a gradual decrease there after. To achieve a perceptually constant scale change, they amend Equations 1 and 2 as follows:

$$speed = \frac{v0}{scale}$$

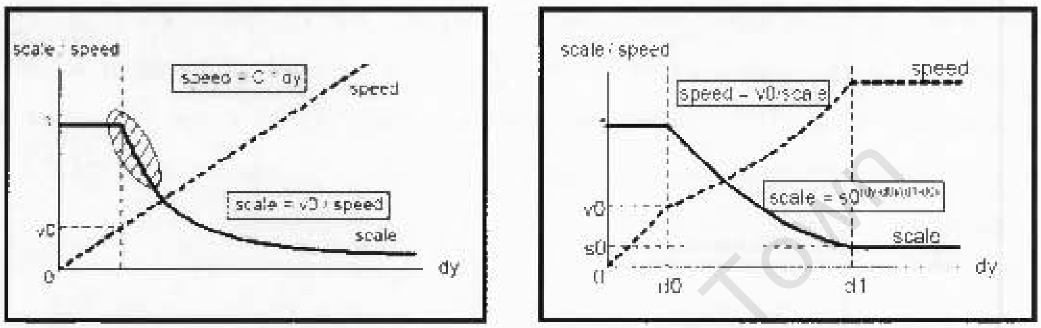
Equation 3

$$scale = s0^{(dy-d0)/(d1-d0)}$$

Equation 4

Figure 3.1b illustrates the behavior of these two equations. Below $d0$ the user can scroll through the information space without any zooming taking place. The information space is also shown at it maximum level of detail. Between $d0$ and $d1$ the scroll speed increases and the zoom level decreases in parallel as defined by Equations 3 and 4. Above $d1$, no further zooming or scrolling takes place. $s0$ specifies the minimal zoom level. $v0$ specifies the scroll speed at $d0$.

Although the graph in Figure 3.1b reflects the desired scrolling and zooming behavior it does not accurately reflect the effect of the equations. While the scale is 1 when the mouse distance equals $d0$, for values of dy greater than zero and less than 1, the scale is greater than 1. Also, neither the scale nor the speed becomes constant at $d1$. Equations 3 and 4 only define the behavior between $d0$ and $d1$. Outside this range the behavior has to be controlled programmatically. These equations also conflict with the notion that zooming is dependent on the scroll speed. In fact, the opposite is true, the scroll speed is dependent on the zoom level. The zoom level is dependent on the displacement of the mouse. Furthermore, Igarashi and Hinckley [49] modified their equations during implementation by introducing a delay in scaling. This was necessary to avoid instantaneous zooming to full size when the mouse was released and undesirable swelling affects when the scrolling direction was reversed.



a b

Figure 3.1: (a) Behaviors of equation 1 & 2 and (b) equations 3 & 4 [49]

Igarashi and Hinckley [49] conducted an informal usability evaluation of two interfaces, a web browser and a map browser. With the web browser, participants were required to locate images using a standard scrolling interface and a SDAZ interface. The task completion times were approximately equal, although six out of seven participants preferred using SDAZ. With the map browser, participants were required to navigate to a marked location using a traditional panning-and-zooming interface and a SDAZ interface. The task completion times were too diverse to provide evidence regarding the interaction modes. In the subjective evaluation, four participants preferred the SDAZ interface and three preferred the traditional panning-and-zooming interface.

Cockburn and Savage [21] conducted a more rigorous experiment evaluation of their own implementation of the SDAZ document and map viewing application. They made use of sophisticated graphical processing techniques to provide more responsive, smoother scroll and zoom animations. Their results were much more promising and showed SDAZ in a new light. In their study, they found that participants were 22% faster when using SDAZ than when using a commercial document viewer. In map browsing, the performance benefits increased to 43%. Furthermore, workload assessments, preferences and participant's comments all amplified the efficiency and effectiveness of the automated zooming approach.

Cockburn and Savage [21] recommend using SDAZ for information spaces of intermediate size. They point out that traditional scrolling is sufficient for small spaces and that query-based search methods are required for large spaces (thousands of screens). Unfortunately, they do not provide dimensions for small, intermediate or large information spaces. They do however state that the experiment was conducted on a 19 inch display with a resolution of 1024 x 768 pixels.

The dimensions of the information space and the resolution are more critical for small screen devices. Take for example an information space of 1024 x 768 pixels, the whole space is viewable on a 19 inch display (with a resolution of 1024 x 768 pixels), but only 9.8% of the space is visible on a HP iPAQ 4100 (with a screen resolution of 320 x 240 pixels) and 4.7% on a Nokia 6680 (with a screen resolution of 176 x 208 pixels). The definitions of small, intermediate and large information spaces must be amended for application on small screen devices. In the next section, some changes are suggested to Igarashi and Hinckley's original equations to take in to account the proportion of the information space that is visible on the screen.

3.4 Extending SDAZ for small displays

The new algorithms that are proposed in this section are loosely based on Igarashi and Hinckley's revised equations. Like Igarashi and Hinckley, the scroll speed is linearly dependant on the current zoom level. In fact, the notion of scroll speed is replaced with two factors which denote the displacement in the x and y dimensions.

The scroll speed is computed with the following algorithm:

```
sdX << 0
sdY << 0
if(no horizontal constraint)    sdX << C * visibleArea * scaleValue * dX
if(no vertical constraint)      sdY << C * visibleArea * scaleValue * dY
if(bounds of information space in X dimension not visible)
    scroll horizontally by sdX
if(bounds of information space in Y dimension not visible)
    scroll vertically by sdY
```

Algorithm 1

sdX and *sdY* are the displacements in the x and y dimension. C is a constant (where C > 0) that is used to control the scroll speed. This value can be customized by users. It can be used to compensate for different device capabilities, such as processing power or hardware graphics support. *visibleArea* is a real value (0 < *visibleArea* <= 1) that indicates the proportion of the information space that is currently visible in the viewport when the information space is displayed at its initial scale. This allows the algorithm to support displays with different screen resolutions. *scaleValue* is a real value (0 < *scaleValue* <= 1) that specifies the current zoom level. When the *scaleValue* is 1 the information space is displayed at its maximum level of detail. When the *scaleValue* is 0.25 the information space is displayed at quarter of its original size. *dx* and *dy* represent the displacement of a mouse in the x and y dimensions and are restricted by the bounds of the information space in each dimension. The units for algorithms 1 and 2 are dependent on the input mechanism that is used. For example, when using a touch screen the units are pixels or when using a tilt-controller the units are degrees.

The *scaleValue* is computed using Algorithm 2. The *dragDist* is the current displacement of a mouse. *minThreshold* specifies the displacement of a mouse below which only scrolling is operational and no zooming takes place. *maxThreshold* represents the displacement of a mouse above which the scroll speed is maintained constant at the maximum scroll speed and the zoom level is fixed at the *minScaleValue*. The *minScaleValue* is a real value (0 <= *minScaleValue* <= 1) that specifies the maximum value at which the information space be zoomed out. The value *A* reflects the proportion of the distance between *minThreshold* and *maxThreshold* at which the mouse is currently located.

```
If( dragDist has changed                          AND
   Full width of information space is not visible    AND
   Full height of information space is not visible   AND
   minThreshold <= dragDist <= maxThreshold          AND
  scaleValue > minScaleValue)
{
     A << abs( (dragDist – minThreshold) / (maxThreshold – minThreshold) )
     scaleValue << minScaleValue – ( (1-A) * (1-minScaleValue) )
}
```

Algorithm 2

The five conditions in Algorithms 2 ensure that:

- Zooming only occurs when there is a change in the displacement of a mouse.
- No further zooming takes place when the full width of the information space is visible.
- No further zooming takes place when the full height of the information is visible.
- Zooming only takes place when the displacement of the mouse is within the allowable bounds.
- No further zooming out takes when the minimum zoom level is reached.

The two algorithms are executed during a pointing action – for example, while a button is depressed on a mouse or while a stylus is in contact with the screen. The behavior of these two algorithms is illustrated in Figure 3.2. SDAZ is initiated when a user presses a mouse button. Scrolling begins when the mouse is moved away from its initial location. The zoom level is adjusted automatically between the *minThreshold* and *maxThreshold* to reduce the high visual flow and consequent blurring that occurs when the information space is scrolled quickly at its normal scale. Within these bounds, the scroll speed is proportional to the displacement of the mouse, while the zoom level is inversely proportional to the displacement of the mouse. When the pointing action ends (for example, when the mouse button is released or the stylus is no longer in contact with the screen) the information space is smoothly animated in to its normal scale.

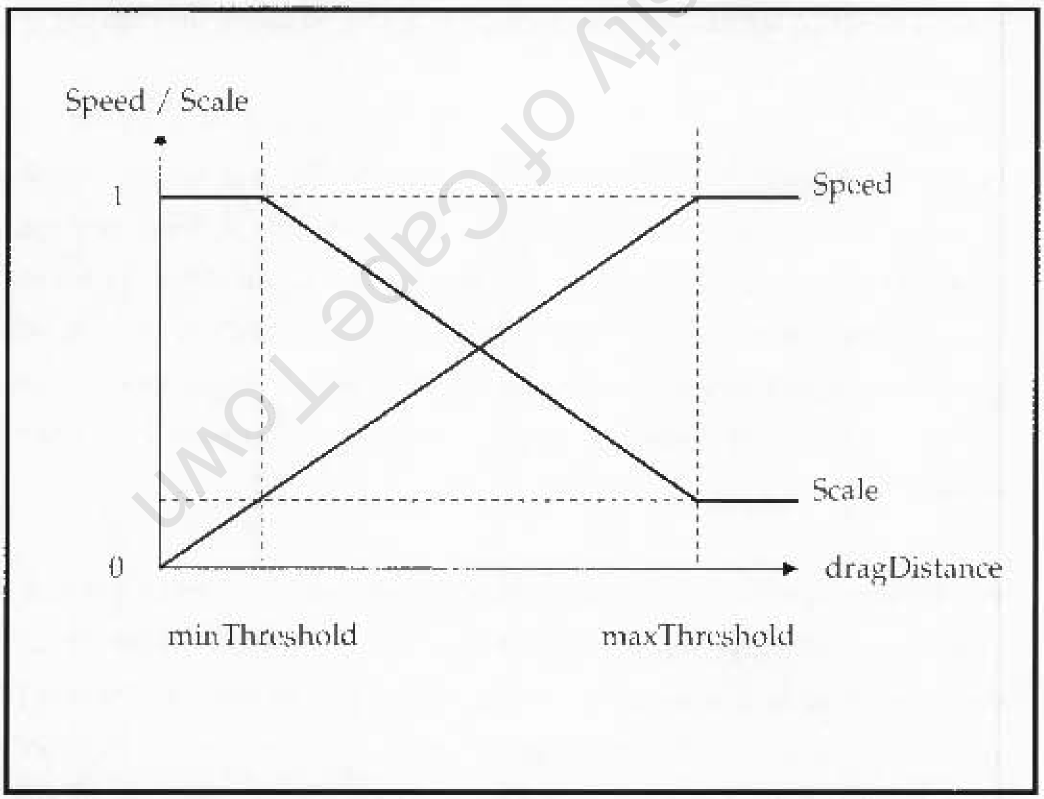


Figure 3.2: Graph illustrating the relationship between the scroll speed and zoom level

These algorithms are generic and can be applied to a number of different applications such as map browsers or document viewers. Consequently, they do not take application specific user needs into account. The next section discusses how SDAZ is tailored to make it more suitable for browsing photographs.

3.5 Visual photo search interfaces

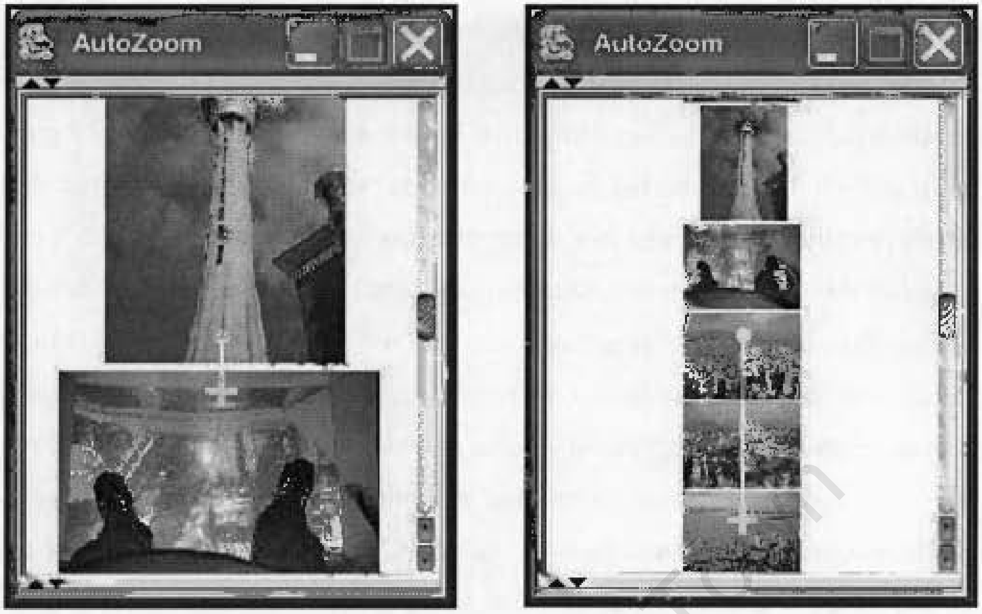
The design of the visual photo search interfaces takes in to account the two core features that are emphasized by users (chronological arrangement and browsable thumbnails [105]) and also the design requirements that are outlined in Section 3.1. The requirements in Section 3.1 are met by basing the design of the visual photo search interfaces on SDAZ, as SDAZ provides dynamic control of the presentation mode and the presentation rate. It is also makes optimal use of the human visual system by maintaining the rate at which users are able to process images. Our version of SDAZ can also be personalized by simply adjusting the constant C (see Algorithm 1).

Three visual photo search interfaces are presented in this section, *AutoZoom*, *ManualZoom* and *DiscreteZoom*. The *AutoZoom* interface is a direct extension of SDAZ. The *ManualZoom* interface is inspired by SDAZ, but differs by providing manual controls for scroll speed and zoom level. The *DiscreteZoom* interface reflects the features found in commercial browsers such as the iPod Photo browser or Pocket ACDSee Photo browser. It was developed for the purposes of comparative evaluation.

For the new interfaces, the photographs are presented in a vertical list that is a single image wide (see Figure 3.3 and Figure 3.5). The motivation for this design decision was based on two studies. The first study by Wang and his colleagues [128] was conducted to find the optimal presentation mode for small display devices. They evaluated three interfaces: thumbnail layout, Flip-zooming layout and an RSVP layout. They found no significances between the three presentation modes in terms of the time taken to locate photographs and the success in completing tasks. Users preferred the RSVP layout as it had the most appropriate photo size for the tasks. The second study by Spence and his colleagues [118] was conducted to identify the best presentation mode for displaying images, static (showing 16 images concurrently), mixed (showing 4 images concurrently) and slideshow (a single image at a time). They found that mixed mode was the least error prone and most preferred as it provided the best comprise between the overview and detail. By displaying the photographs in a linear structure and supporting smooth transitions between the overview and detail, users can easily select the level of detail that is appropriate for a task. Users can zoom into an individual photograph to inspect detail and smoothly zoom out to obtain an overview of the event. The linear structure makes it easier to read events, as events naturally occur in sequence. The ability to zoom out and see a number of events in sequence provides users with additional contextual cues (such as subtle changes in color or shape) that allow events to be identified even when individual photographs are too small to be recognized. The linear structure with added zooming capability gives users have the best features of a thumbnail browser, coupled with more flexible control of the presentation mode. The photographs are arranged chronologically, with the most recent photograph being placed at the top of the list.

3.5.1 *AutoZoom* interface

The *AutoZoom* interface (see Figure 3.3) is operated by vertical dragging actions with the pointing device. These actions control the rate at which images scroll through the viewport, the image size (zoom level) and the scroll direction. The vertical center of the viewport acts as the threshold for direction change—dragging above the center moves the images downwards and vice versa. Image size is inversely proportional to the distance of the pointer from the vertical center, and changes dynamically as the pointer moves either away from or towards the center. The perceived effect to the user is the faster you go, the higher you fly. Images are not reduced beyond a minimum (user specified) size threshold. Once this threshold is reached, an acceleration function maps further increases in the drag distance proportionally to the scroll speed (see Figure 3.4). When the user completes an action by releasing the pointing device, the images are smoothly animated back to their normal size at the current location in the list.



(a) (b)

Figure 3.3: The *AutoZoom* interface: as cursor is dragged away from center, scroll speed/image size change. (a) moderate speed, images slightly reduced; (b) fast speed and minimum size (Cross added for clarity)

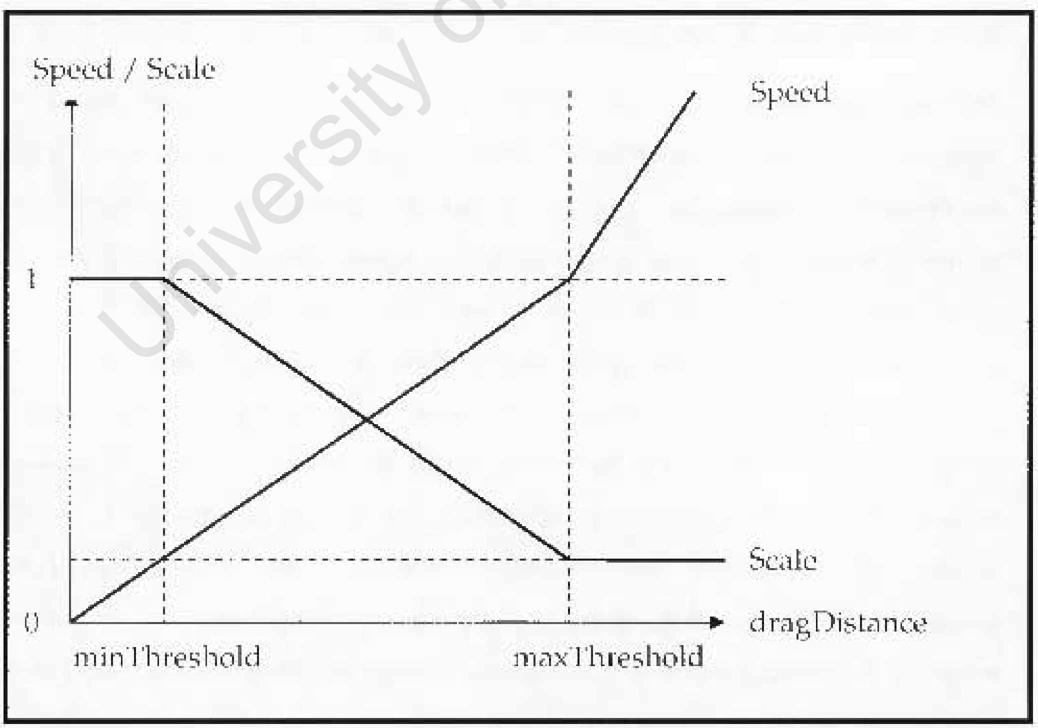


Figure 3.4: Graph showing the behavior of the *AutoZoom* interface

To achieve this behavior Algorithm 1 is modified as follows:

```
sdX << 0,
sdY << 0

if(scaleValue > minScaleValue)
{
        if(no horizontal constraint)
                sdX << C * visibleArea * scaleValue * dX
        if(no vertical constraint)
                sdY << C * visibleArea * scaleValue * dY
}
else
{
        if(no horizontal constraint)
                sdX << K * C * visibleArea * minScaleValue * dX
        if(no vertical constraint)
                sdY << K * C * visibleArea * minScaleValue * dY
}

if( no horizontal constraint  AND
   bounds of information space in X dimension not  visible)
{
        scroll horizontally by sdX
}
if( no vertical constraint AND
   bounds of information space in Y dimension not  visible)
{
        scroll vertically by sdY
}
```

Algorithm 3

The constant K (where $K > 0$) is used to control the scroll speed once the minimum zoom level is reached. The acceleration function allows users to quickly spurt between events, moving rapidly over irrelevant photographs and slowly over more relevant ones.

The *scaleValue* is computed using the following algorithm:

```
If( dragDist has changed                                AND
    minThreshold <= dragDist <= maxThreshold             AND
    scaleValue > minScaleValue                           AND
    ((no horizontal constraint                           AND
    Full width of information space is not visible)      OR
    ((no vertical constraint                             AND
    Full height of information space is not visible)))
{
     A << abs( (dragDist – minThreshold) / (maxThreshold – minThreshold) )
     scaleValue << minScaleValue + ( (1-A) * (1-minScaleValue) )
}
```

Algorithm 4

Algorithm 2 is amended to allow the images to be scaled smaller than dimensions of the screen. Scaling is allowed if the full length of the unconstrained dimension of the information space is not visible and if the *scaleValue* is greater than the *minScalevalue*.

3.5.2 *ManualZoom* interface

The *ManualZoom* interface provides independent controls for adjusting the presentation mode and the presentation rate. At any level of detail, users can select a suitable scroll speed. This supports users more adequately in performing attentional selection when searching through images (see Chapter 2.4.2). The *ManualZoom* interface also allows the scroll speed and the zoom level to be controlled in parallel using a single input device. This provides the advantages of independent controls without the cognitive overhead of simultaneously manipulating two input mechanisms. The ability to zoom without scrolling allows users to obtain an overview without having to scroll away from a point of interest.

(a) (b)

Figure 3.5: The *ManualZoom* interface: (a) moderate scroll speed and small image reduction; (b) maximum speed and minimum size. (Cross added for clarity)

For the *ManualZoom* interface, vertical drag operations control the scroll speed and the direction as with the *AutoZoom* interface, but do not control the image size (or zoom level). The zoom level is controlled by horizontal movement of the pointing device away from the horizontal center of the viewport to the right-hand or left-hand side of the display. The zoom level is inversely proportional to the horizontal drag distance. Figure 3.5a shows the pointers position (indicated by the cross), leading to a moderate scroll speed with small image reduction: the user is dragging below and slightly to the right of the viewport center. In Figure 3.5b, the user has dragged the pointer to the right-hand corner of the display, producing the maximum scrolling speed and the minimum image size. Returning the pointer to the center of the viewport returns the images to the full size. As with *AutoZoom* interface, when the user releases the pointer the images smoothly animate back to their normal size.

The scroll speed is computed using the following algorithm:

```
sdX << 0,
sdY << 0
A << K * C * visibleArea * minScaleValue

if(no horizontal constraint)
      sdX << S * A * dX

if(no vertical constraint)
     sdY << S * A * dY

if( no horizontal constraint AND
    bounds of information space in X dimension not visible)
      scroll horizontally by sdX

if( no vertical constraint AND
    bounds of information space in Y dimension not visible)
      scroll vertically by sdY
```

Algorithm 5

This algorithm ensures that the maximum scroll speed for the *AutoZoom* and *ManualZoom* interfaces is the same. This is necessary for comparative purposes. The constant S (where $S > 0$) is be used to control the scroll speed. When $S = 1$, the maximum scroll speed, is equivalent to the *AutoZoom* interface.

The scale is computed using the following algorithm:

```
If( dragDist has changed                                AND
    dragDist <= maxDragDistance                         AND
    scaleValue > minScaleValue                          AND
    ((no horizontal constraint                          AND
    Full width of information space is not visible)     OR
    ((no vertical constraint                            AND
    Full height of information space is not visible)))
{
        A << abs( (dragDist) / (maxDragDistance) )
        scaleValue << minScaleValue + ( (1-A) * (1-minScaleValue) )
}
```

Algorithm 6

maxDragDistance represents the drag distance above which no further zooming takes place. As stated above, both Algorithms 5 and 6 are computed in parallel.

(a) (b)

Figure 3.6: The *DiscreteZoom* interface: (a) the thumbnail view and (b) the enlarged view

3.5.3 *DiscreteZoom* interface

The *DiscreteZoom* interface (see Figure 3.6) presents photographs in a 4x3 row-and-column scrollable list ordered by creation time. Users can click/tap on the desired photo to view an enlarged version. The selected photo is animated to fill the screen. Similarly, users can click on the enlarged photo to return to the thumbnail view

3.5.4 Personalization

The settings for each interface can be customized as shown in Figure 3.7. With the *AutoZoom* interface users are able to customize the scroll speed, the minimum thumbnail size, the minimum threshold and the maximum threshold. The minimum and maximum threshold govern the range in which zooming occurs. The maximum threshold specifies the point at which the minimum thumbnail size is reached. For the *ManualZoom* interface users can similarly customize the scroll speed and minimum thumbnail size. The *DiscreteZoom* interface allows users to specify the number of columns of thumbnails to show within the view. All controls allow users to dynamically adjust these values.

(a) (b) (c)

Figure 3.7: Customization panel for each search interface: (a) *AutoZoom* interface, (b) *ManualZoom* interface and (c) *DiscreteZoom* interface

All three interfaces have a scrollbar to provide context. The scrollbar can be used to quickly find an approximate location in the photo collection. For the *AutoZoom* and *ManualZoom* interface, as the user begins to drag the slider the images are immediately reduced to their minimum size and normal scrolling follows. At the end of the scrolling operation the images are expanded to their normal size. Hence, when scrolling, users have an overview of the image set.

3.6 Experimental Evaluation

The objective of the experiment was to compare the three visual photo search interfaces in terms of user performance and subjective preferences.

3.6.1 Hypothesis

The hypotheses were as follows:

1. Both the *AutoZoom* and *ManualZoom* interfaces support faster navigation to target photographs than the *DiscreteZoom* interface.

2. Both the *AutoZoom* and *ManualZoom* interfaces support more accurate identification of target photographs than the *DiscreteZoom* interface.

3. Subjective task load levels are lower for both the *AutoZoom* and *ManualZoom* interfaces than the *DiscreteZoom* interface.

3.6.2 Participants

Seventy-two participants (38 male and 34 female) took part in the experiment. Sixty-one participants were students (either postgraduate or undergraduate), 6 were lecturers and 5 were software developers. 45 of the participants had previously used photo management software, but only 5 on a small screen device. None of the participants had used SDAZ interfaces. 70 participants described themselves as casual photographers (i.e. occasionally take photographs). Two participants described themselves as professional photographers (e.g. take photos for magazines or weddings).

Participants were recruited through poster and e-mail advertisements and each received a minor gift for their voluntary participation.

3.6.3 Method

A between-subject design was employed (see Table 3.1 and Table 3.2). Seventy-two participants were randomly allocated to one of three groups, each containing 24 participants. Each group used one of the three interface designs to complete photo navigation tasks. Random allocation was used to avoid any external bias that might be introduced by manipulating the participants in each group. Seventy-two participants were needed for a complete counter balanced experimental design of all the factor level combinations. By having a large number of participants, we were able to ensure that variables such as age, gender, and experience were fairly evenly distributed across each group (see Table 3.3).

The independent variables were as follows:

- *Interface*: Between-subjects variable with three levels: *AutoZoom*, *ManualZoom* and *DiscreteZoom*;

- *Task type*: The task type was a within-subjects variable with three levels: *Event* (participants searched for a set of photos relating to a particular well-defined event, e.g., "locate the Motor Rally"); *Single* (participants searched for an individual photo containing a specified feature, e.g., "Find this image of the Sky Tower"); and *Property* (participants searched for a set of photos taken at different events, but all sharing a property, such as all the photos containing an specific object, e.g., "Count all the photos that contain an hot-air balloon");

- *Navigation distance*: For *Event* and *Single* task types only. Within-subjects variable with two levels: short and long. Short distances were no more than half the length of the photograph list, and long distances were always more than half the length.

		Interface Type					
		AZ		M Z		D Z	
		Navigation Distance					
		short	long	short	long	short	long
Event Or Single	small	G_1	G_1	G_2	G_2	G_3	G_3
	large	G_1	G_1	G_2	G_2	G_3	G_3

Table 3.1: A 3×2×2 repeated measure factorial design for *Event* and *Single tasks*.

	Interface Type		
	AZ	MZ	DZ
Property	G_1	G_2	G_3

Table 3.2: 3x1 repeated measure factorial design for *Property* task.

		Interface type (number of participants)		
		AZ	**MZ**	**DZ**
Participant attributes	**Male**	13	11	14
	Female	10	12	12
	Photo management software experience on desktop	15	14	15
	Photo management software experience on PDA's	1	2	2
	Casual photographers	23	24	23
	Professional photographer	1	0	1
	Students	22	18	21
	Lecturers	2	2	2
	Software developers	2	1	2

Table 3.3: Participant demographics.

Events could be small (3 or fewer photos), or large (more than 3 photos). A photograph feature could also be small or large. The area of a small feature (e.g. a small child in a forest scene) was 1/8th (or less) of the total area of an image, while the area of a large feature was greater than 1/8th of the total area of an image (e.g. a skyscraper). In both categories we chose pictures that maximized the amount of space available, so the area for a small feature was as close to 1/8th as we could get and large features took up most of the space in a photograph.

Each participant completed a total of 27 experimental tasks, using one of the interfaces. For the *Event* task type they completed 3 tasks for each of the 4 navigation distance/event size combinations. For the *Single* task type they completed 3 tasks for each of the 4 navigation distance/feature size combinations. For the *Property* task type they completed 3 tasks (requiring the user to find 16, 30 and 120 images respectively).

The presentation order of the tasks types was counterbalanced to minimize learning effects. For each task-type/interface combination, the presentation order of the factor level combinations was also counter balanced to minimize learning effects

3.6.4 Materials

The experiment was carried out on a desktop computer with a 1.7GHz processor, 1GB of RAM, and running Microsoft Windows XP. The viewport size for all three interfaces was set to 240x340 pixels to simulate the display of the HP h5550 Pocket PC. A mouse was used as a stylus surrogate.

This experiment was simulated on a desktop computer because when the software was written, devices such as the HP Pocket PC did not have sufficient processing power and memory to run these applications. In fact, other studies in the field have also simulated experiments in the same way [29][118]. Regardless of this fact, the key factor for investigation in this study is the screen size rather than the interaction device, hence the focus on adapting SDAZ for small screen devices. Of course it is important to evaluate these interfaces on actual devices. This is done in Chapters 4 and 6.

A single set of 300 of photographs was used to provide a consistent set of stimuli across all tasks, participants and conditions. This was done to ensure that we could easily isolate the effects that were due to the other independent variables and also reduce the number of independent variables. For similar reasons, other similar experiments have also used a standard set of pictures [29][118][128]. A major concern with using a stock collection is that it calls in to question the ecological validity of the experiment. To minimize this, the stock collection was an actual set of pictures taken on from a trip to New Zealand. Another concern is that it introduces another variable, the participant's ability to learn a set of photographs. We conducted extensive training sessions to familiarize the participants with the set of photographs and minimize learning effects when conducting the experiments. It can also be argued that as participants have varying cognitive abilities, the ability to recall their own personal photographs is likely to differ between them anyway.

For the experiment, the *DiscreteZoom* interface was set to show 12 thumbnails on the screen. The minimum thumbnail size for the *AutoZoom* and *ManualZoom* interface was set to the same size as the *DiscreteZoom* interface. For the *AutoZoom* and *ManualZoom* interface the constant K was set to 4 and the constant C was set to 10 to provide a responsive and natural interaction. The maximum scroll speed was identical for both the *AutoZoom* and *ManualZoom* interfaces. The participants were not able to changes these settings during the experiment.

A printed questionnaire was developed and issued to participants to record their background and experiences. Additional software was developed for training users and managing the experiment.

3.6.5 Procedure

On arrival, participants were asked to read a summary of the experiment and provide consent to continue if they were in agreement (see Appendix A for experimental materials). They then completed a profile questionnaire and were given 15 minutes to familiarize themselves with the set of photographs to be used in the experiment (see Figure 3.8). Following this, they were asked to read instructions that provided a detailed description of each task type and also explained the operation of their assigned interface.

Figure 3.8: Photo collection training website

The operation of the interface was then demonstrated and the participants were given 10 minutes to explore the operation of the software for themselves. Following this, they were given a set of training tasks of the same form as the experimental tasks. Participants were encouraged to ask questions throughout the training period. Once the training tasks were completed, participants could take a short break before commencing the experimental tasks. One aim of the training session was to allow users to familiarize themselves with the image set so that any learning effects during the experimental tasks would be reduced.

Participants controlled progress of the experimental session via an on-screen dialog that allowed them to initiate a task, displayed task instructions, and enabled them to indicate when a task was completed (see Figure 3.9). At the start of every task, the viewport was reset to show the beginning of the image list.

Event tasks were described textually. An event was found by selecting any one of the photographs within the event. For *Single* tasks, participants were shown the target photograph and its corresponding caption. For both *Event* and *Single* tasks, users were prompted by the system to retry if their selection was incorrect; they were able to attempt the task as many times as they needed.

Figure 3.9: Task management software

For *Property* tasks, participants were required to count the number of photographs that shared a common property. They were given a field into which to enter the number. On completion of all the tasks participants were requested to fill-out a questionnaire that captured their subjective views of the software.

3.6.6 Data captured

For each task the software automatically recorded a range of events including: the time to complete tasks; distinct scrollbar operations and distinct zoom operations.

For *Property* tasks there were a target number of photos (A); in completing the task, a user found a number of images (C). Accuracy was then calculated as follows:

$$Accuracy = 100\left(1 - \frac{|A - C|}{A}\right)$$

This measure penalizes overlooked images and false positives. Finally, we collected subjective responses about the workload required to complete tasks, as measured by the NASA task load index [45]. Reponses to six measures (i.e. mental demand, physical demand, temporal demand, effort, performance, frustration level) were captured using a five point Likert scale, with lower values reflecting lower task loads. During the experiment, critical events and comments were also noted.

3.6.7 Results

In all cases, the statistical data was subjected to significance testing using the analysis of variance method (ANOVA). Before conducting any ANOVA's we tested the assumptions of normality and homogeneity of variances to ensure they were not violated.

	Interface Type					
	AZ		**MZ**		**DZ**	
	mean	s.d.	mean	s.d.	mean	s.d
Time (seconds)	35.20	40.70	36.68	40.26	44.10	48.51

Table 3.4: Search times for the three interface types.

Over all tasks the interface type had no significant effect on the task completion times ($F(2, 645)=2.6192$, $p=0.07364$). However, the *AutoZoom* interface was significantly faster at performing tasks than the *DiscreteZoom* interface ($F(1, 430)=4.2706$, $p=0.03938$), with means of 35.20 seconds and 44.10 seconds, respectively (see Table 3.4). Overall, there was no significant difference in terms of the time taken to complete tasks between males and females. There were no other significant differences.

Locating *Events*

The *AutoZoom* and *ManualZoom* interfaces were significantly faster than the *DiscreteZoom* interface when searching for small events (F(2,69) = 5.0597, p=0.00890), with means of 26.01 seconds, 29.41 seconds and 45.46 seconds, respectively (see Table 3.5). Over all *Event* tasks, though, the interface type had no significant effect on the task completion time (F(2,69)=1.2848, p=0.28323).

		Interface Type					
		AZ		**MZ**		**DZ**	
		Mean Search Time (seconds)					
		mean	s.d.	mean	s.d.	mean	s.d
Navigation Distance	Short	25.97	26.34	30.62	38.91	43.15	51.49
	Long	20.93	23.66	27.45	26.56	21.99	17.30
Event Size	Small	**26.01**	**28.15**	**29.41**	**31.16**	**45.46**	**50.32**
	Large	20.89	21.47	28.66	35.40	19.68	17.55

Table 3.5: Mean completion times for each factor level combination for *Event* tasks.

Regardless of interface, participants took significantly longer to locate events which were a short distance away (F(1,69) = 8.9667, p=0.00381), with short and long distance means of 33.25 and 23.46, respectively. At long navigation distances, large events were found significantly faster than the small events (F(1,69)=6.5946, p=0.01240), with mean search times of 14.04 seconds and 32.88 seconds. For short navigation distances, the event size had no significant interaction with the amount of time taken to locate an event, with search means of 34.38 seconds (large events) and 32.12 seconds (small events).

Locating *Single* photographs

The *AutoZoom* interface was significantly faster at finding single photographs than the *DiscreteZoom* interface at long navigation distances (F(1,46) = 9.5749, p=0.00335) with means of 28.90 seconds and 44.06 seconds, respectively (see Table 3.6). Both the *AutoZoom* and *ManualZoom* interfaces were significantly faster than the *DiscreteZoom* interface when searching for images with small features ($F(2,69) = 3.1596, p = 0.04865$) with means of 39.15 seconds, 34.52 seconds and 48.69 seconds, respectively. Over all *Single* tasks, though, the interface type had no significant effect on task completion times ($F(2,69)=0.79012$, $p=0.45785$).

		Interface Type					
		AZ		MZ		DZ	
		Mean Search Time (seconds)					
		mean	s.d.	mean	s.d.	mean	s.d
Navigation Distance	Short	30.76	37.38	24.47	19.20	25.32	19.41
	Long	**28.90**	**20.76**	31.96	25.93	**44.06**	**37.36**
Feature Size	Small	**39.15**	**37.34**	**34.52**	**25.83**	**48.69**	**35.69**
	Large	20.52	16.08	21.91	17.93	20.67	16.65

Table 3.6: Mean completion times for each factor level combination for *Single* tasks.

Regardless of interface, participants took significantly less time to locate single images that were a short distance away ($F(1,69)=11.330$, $p=0.00125$), with short and long distance means of 26.85 seconds and 34.98 seconds respectively. Also, images with smaller features took significantly longer to detect than those with larger ones ($F(1, 69)=61.446$, $p=0.00001$), with small and large means of 40.79 seconds and 21.04 seconds respectively.

Locating photographs with a *Property*

The *AutoZoom* and *ManualZoom* interfaces were significantly more accurate than the *DiscreteZoom* interface ($F(2,69)=14.614, p=0.0001$), with mean accuracy rates of 92.38%, 89.98% and 76.15%, respectively.

Over all *Property* tasks (see Table 3.7), the interface type had no significant effect on the task completion time ($F(2,69)=1.5150$, $p=0.22704$), with search means of 103.66

seconds (*AutoZoom*), 101.14 seconds (*ManualZoom*) and 127.89 seconds (*DiscreteZoom*).

	Interface Type					
	AZ		**MZ**		**DZ**	
	mean	s.d.	mean	s.d.	mean	s.d
Time (seconds)	103.66	59.74	101.14	59.50	127.89	57.08

Table 3.7: Mean completion times for *Property* tasks.

Subjective Preference

There was a significant difference between the mean task load ratings for the three interfaces ($F(2,69) = 6.0275$, $p=0.00387$), with mean ratings of 2.31 (*AutoZoom*), 2.53 (*ManualZoom*) and 3.01 (*DiscreteZoom*).

	Interface Type					
	AZ		MZ		DZ	
	mean	s.d.	mean	s.d.	mean	s.d
Mental	**2.67**	**0.92**	3.00	1.06	**3.38**	**0.77**
Physical	2.04	1.16	2.46	1.50	2.88	1.19
Temporal	2.95	1.20	2.96	1.20	3.50	0.93
Performance	1.77	0.83	1.88	1.17	2.13	0.89
Effort	**1.75**	**0.90**	2.25	1.11	**2.58**	**1.14**
Frustration	**2.67**	**0.92**	**2.63**	**0.92**	**3.63**	**1.10**
Average	**2.31**	**0.64**	**2.53**	**0.84**	***3.01***	**0.67**

Table 3.8: NASA task load rating, lower values reflect lower workload.

Looking at the individual factors measured by the task load index (see Table 3.8), participants found both new interfaces significantly less frustrating than the *DiscreteZoom* interface $F(2,69) = 7.9593$, $p= 0.00078$). Furthermore the mental workload ($F(1,46) = 8.4033$, $p = 0.00572$) and effort ($F(1,46) = 7.9310$, 0.00713) were significantly lower for the for the *AutoZoom* interface than the *DiscreteZoom* interface.

3.6.8 Discussion

Considering the results in the light of the three hypotheses noted in Section 3.6.1

1. ***Both the AutoZoom and ManualZoom interfaces support faster navigation to target photographs than the DiscreteZoom interface.*** The results indicate the new techniques performed as well and in some cases better than the *DiscreteZoom* interface. Overall the *AutoZoom* interface was significantly faster at performing tasks than the *DiscreteZoom* interface. More specifically, both new interfaces were significantly faster when finding *Single* photos containing small-sized features as well as detecting *Events* consisting of a small number of photos. The *AutoZoom* interface was also significantly faster than the *DiscreteZoom* interface at locating *Single* images at long navigation distances.

2. ***Both the AutoZoom and ManualZoom interfaces support more accurate identification of target photographs than the DiscreteZoom interface.*** The new techniques were significantly more accurate when finding a set of photographs that fit a given description.

3. ***Subjective task load levels were lower for both the AutoZoom and ManualZoom interfaces than the DiscreteZoom interface.*** The results of the task load calculations show that participants perceived the new systems to be significantly less onerous than the *DiscreteZoom* interface.

It is worth remembering that none of the participants had previous experience of SDAZ-type interfaces, while all were familiar with the conventional scrolling approach of the *DiscreteZoom* interface. It is encouraging, then, to see such consistently good performance with the new schemes after minimal training. During task completion, the average amount of time spent operating the zoom/scroll control with the new interfaces was 22.5 seconds; this is nearly four times the duration spent using the scrollbar (5.9s). We are satisfied, then, that the benefits provided by the new interfaces come from the integration of scrolling and zooming.

Small features in an image, small groups of photographs and individual target photos are more easily overlooked with the *DiscreteZoom* interface, as they scroll past at thumbnail size; the explicit zoom-in/zoom-out operations needed to check individual image contents also contributes to the slower performances. Such problems with grid-based thumbnail browsing have been recognized by others who suggest, for example, processing the images to present only the salient details [129].

One counterintuitive result seen in relation to the *Event* tasks was that it took longer to find events close to the starting point than those which were further away. On closer inspection of the data, this effect was largely due to a small event that was in between two large events that had similar types of pictures. This event was also near the top of the collection. We observed that the participants would begin searching rapidly with scant regard for the initial images they were scrolling past; there was then ample opportunity for participants to easily overshoot early targets. On possible explanation is that participants could have scrolled right passed the event, thinking it was part of an event. Participants then spent a significant amount of time trying to locate the event between these two events. Further evidence of this is the fact that same effect was observed for each of the interfaces.

3.7 Summary

This section summarizes the chapter and provides an outlook to subsequent chapters. Section 3.1 describes design requirements for visual photo search tools. It also motivates an evolutionary design approach, rather than revolutionary one. Section 3.2 provides a critical analysis of methods for navigating large information spaces. SDAZ is identified as a suitable approach, due to its ability to meet the design requirements that specified in Section 3.1. Section 3.3 discusses the mechanics of the SDAZ technique. Section 3.4 discusses algorithmic extensions that are made to SDAZ to make it more suitable for small displays. In particular it takes in to consideration the proportion of the information space is that visible on a low resolution screen. Section 3.5 discusses how this technique is used to create two novel visual photo search techniques. It also discusses the design and development of a traditional-grid interface that is used for comparative purposes in the user study. Section 3.6 compares the three visual photo search interfaces in terms of user performance and subjective preferences. Lastly, Section 3.6 reflects on the main findings of this chapter in terms of the overall thesis goal.

The aim of this chapter was to develop a photo search tool that addresses the limitations of current visual photo search techniques. The findings in the experiment show that the *AutoZoom* and *ManualZoom* interfaces offer substantial improvements over current visual photo search tools. The real benefit of these new techniques is they can easily be extended. While the new techniques have been implemented to accommodate a device using a pointer (e.g. a stylus), they can be extended for use with other interaction devices. For example, the *AutoZoom* interface could be used with physical dial-type wheels as seen on the iPod or the *smartPad* (proposed by Rekimoto, for use in mobile phones [102]) providing one-handed interaction. Meanwhile, the *ManualZoom* interface could permit the use of joystick-type mechanisms. Furthermore, when the photographs are zoomed out the white space on either side of the photo strip allows additional contextual cues such temporal data or folder information, both of which are used extensively when searching for photographs [62]. The next chapter begins by discussing how the *AutoZoom* and *ManualZoom* interfaces are modified to include this functionality. It goes on to use this search tool as platform from which to ask more interesting questions about photo search.

Chapter 4

Understanding how to support search tasks

4.1 Introduction

In the previous chapter, the *AutoZoom* and *ManualZoom* interfaces were designed to address the limitations with current visual photo search tools on small display devices. The goal of the study was to gauge whether these new techniques provide a better search experience in terms of specific criteria such as speed of retrieval, accuracy in identifying photographs and the subjective workload. Although, the results show that the new techniques offer some improvements, they do not provide any insights on how improve the techniques further. What they do provide is an empirically grounded 'platform' from which to ask more questions about search. Questions such as: how do people search for *events*, *singles* and *properties* on small display devices? What problems (or limitations) arise when using the *AutoZoom* and *ManualZoom* interfaces to perform these tasks? Are there any instances where one interface is preferable to the other? Given that folder labels and dates are used extensively when searching for photographs [62], how might we extend these techniques to incorporate this information and what impact might this have when searching for *events*, *singles* and *properties*?

This chapter tackles these questions. The *AutoZoom* and *ManualZoom* interfaces are modified by adding folder labels and temporal data. Both interfaces are evaluated in terms of their ability to locate *events*, *singles* and *properties*. The goal is to identify problems with the interfaces so that they can be addressed to provide a better search experience when performing the three search tasks.

In contrast with the previous chapter, the study is conducted on a small screen device using personal photo collections. A qualitative research paradigm is used to help us understand and develop explanations for questions above. John Creswell

[23] notes that a qualitative research paradigm is beneficial when: the concept is 'immature' because of a lack of previous research and theory; a need exists to explore and describe the phenomenon and to develop theory; and the nature of the phenomenon may not be suited to quantitative measure.

4.1.1 Contributions

The main contributions include:

- *Algorithms to support the AutoZoom and ManualZoom interfaces in limited environments.* The algorithms tackle two major issues with mobile devices: varied input mechanisms, and limited device capabilities. A mapping function is used to provide a unified way of accommodating multiple input mechanisms. A number of optimizations are presented on how to implement these techniques in a processor and memory efficient manner, ensuring that they scale elegantly and are independent of the photo collection size. This is important because future devices will be able to store significantly larger photo collections.

- *Improving our understanding of how people search for events, singles and properties.* Based on an observational study, we found that when searching a photo collection the participants would primarily think of their photographs in terms of events. This seemed to occur irrespective of the task type. For example, when locating a *single* they would first try to associate it with an event and then locate the target photograph. When locating *properties*, they would think of events that were likely to contain target photographs and then navigate from one to the next. Often participants would automatically think of the date as secondary process to identifying a target event. When the date was precisely known, they would navigate directly to the date. Otherwise, they would narrow the search down to a year or month by thinking of when it happened relative to neighboring events or landmark events. The task of locating an event was made easier by the fact that people organize photographs by events. Participants noted that the metadata that is associated with an event encodes the kind of information that they are likely to remember when searching for an event. We observed that this metadata was used extensively when locating target events. Some participants noted that they preferred to use the folder labels to distinguish events rather than scanning through the photographs. We also observed that some participants would organize events according to special themed categories. When searching for an event in a category, the participants would first identify the event, then think of an appropriate category and finally navigate to the category. When the spatial location was known and participants could not pinpoint an event temporally, we found that participants preferred to browse spatially rather than temporally as they felt it was more cognitively demanding to constantly resolve and reevaluate where you are temporally in relation to a target event. Based on these observations it is evident that multiple search strategies are used to locate events.

4.1.2 Outline

Section 4.2 discusses the prototype implementation. It begins by discussing the challenges in porting the *AutoZoom* and *ManualZoom* interfaces to mobile devices. It also discusses some interface changes. Section 4.3 describes the experimental goals and setup. Section 4.4 presents the major findings in the study. Section 4.5 discusses these findings. Section 4.6 reviews the chapter and provides an outlook for subsequent chapters.

4.2 Search tool implementation

Mobile devices have many limitations that must be considered when designing visualization or navigation techniques.

- Limited screen size
- Different input mechanisms: stylus, DPAD, mini joysticks, dial-type wheels
- Limited device capabilities: processing power, memory, storage space

Although device capabilities will continue to improve, future devices will store significantly larger photo collections, requiring scalable and efficient solutions that are independent of the photo collection size. The continuing trend towards miniaturization (Motorola's ultra thin handset, RAZR was the best selling phone in Europe in 2006[1]) will mean that some of these limitations are unlikely to disappear in the near future [20].

The limited screen size was the topic of discussion in the previous chapter. This chapter focuses on addressing issues related to input mechanisms and limited device capabilities. Section 4.2.1 describes how different input techniques are accommodated in a unified way. Section 4.2.2 discusses a number of optimizations for achieving responsive and smooth animation in limited environments. Lastly, Section 4.2.3 discusses some interface modifications that are made to incorporate contextual data.

4.2.1 Input mechanisms

There are four key challenges to consider when supporting multiple input mechanisms. Firstly, the *AutoZoom* and *ManualZoom* interfaces must be designed to accommodate a range of different input mechanisms which can also be manipulated in different ways – for example, pointing-and-tapping with a stylus, one-handed thumb input with a mini joystick, pressing buttons with a D-pad, or tilting a tilt-controlled device. The challenge is mapping these different actions to the scroll speed and zoom level. Secondly, input mechanisms have different resolutions. For example, a four-way joystick or D-pad can only distinguish between two levels (on or off) in any direction, whereas a stylus has a much higher resolution and can distinguish between multiple levels. The challenge is mapping variable scroll speeds and zoom levels on to input mechanisms with very low resolutions. Thirdly, the *AutoZoom* and *ManualZoom* interfaces must support single and dual-handed interaction. When users can pay attention to an interface (such as when they have dead time to fill) dual-handed interaction is likely to be more appropriate. On the other hand, when users are engaged in multiple tasks that require the use of their hands, single-handed interaction might be more appropriate. Fourthly, a control

[1] http://www.telephia.com/html/insights_032906.html

widget must be designed to provide feedback, minimize the screen real-estate for control and avoid distracting the user.

Accommodating multiple input mechanisms

The *AutoZoom* and *ManualZoom* interfaces can be controlled by adjusting one or two values, respectively. This simplicity is the key to supporting multiple input mechanisms. A mapping function is used to convert any input mechanism actions (such as, the displacement of a pointing device such as a stylus, or the duration a button is depressed with a D-pad or the tilt angle with a tilt control device) into translations in the x or y dimension. These translations are then used to determine the zoom level and scroll speed (see Algorithm 1).

maxDistX and *maxDistY* are constants that specify the maximum displacement of an input device. For example, when using a tilt control device, these values would be set to the maximum resolution of the device, such as 360 degrees. The maximum viewable angle can then be determined by assigning appropriate values to the following constants *minThresholdPercentageX*, *maxThresholdPercentageX*, *minThresholdPercentageY*, and *maxThresholdPercentageY*. Some typical values might be 2%, 25%, 2%, and 25%, respectively. This would result in the *minThresholdX*, *maxThresholdX*, *minThresholdY*, and *maxThresholdY* being set to 3.6 degrees, 45 degrees, 3.6 degrees and 45 degrees.

These threshold values determine the boundaries of the positive and negative viewable angle. The positive viewable angle region would be between 3.6 and 45 degrees and the negative viewable angle region would be between -3.6 degrees and -45 degrees. Tilting within either region determines the scrolling direction and scrolling speed. No scrolling occurs between 3.6 and -3.6. This provides a stationary position. For a tilt control device this is particularly useful as it difficult to hold a device at a specific angle. *initValueX* and *initValueY* can be used to specify the stationary position. The value might be set to 45 degrees, so that the device can be held in a more natural position. *currentValueX* and *currentValueY* specify the current tilt angle. *distX* and *distX* specify the displacement from the stationary position. These values are then used to compute the displacement in the x and y dimension.

```
maxThresholdX << (maxDistX*maxThresholdPercentageX) / 2
minThresholdX <<  (maxDistX*minThresholdPercentageX) / 2
maxThresholdY << (maxDistY*maxThresholdPercentageY) / 2
minThresholdY <<  (maxDistY*minThresholdPercentageY) / 2
.....
dX << 0, dY << 0, distX << 0, distY << 0
distX << currentValueX – initValueX
distY << currentValueY - initValueY

if(distX has changed   AND  minThresholdX <= abs(distX)  <= maxThresholdX)
{
    if(distX < 0)
        dX<<(distX + minThresholdX)/(maxThresholdX–minThresholdX)
    else
        dX<<(distX - minThresholdX)/(maxThresholdX–minThresholdX)
}

if(distY has changed   AND  minThresholdY <= abs(distY)  <= maxThresholdY)
{
    if(distY < 0)
        dY<<(distY + minThresholdY)/(maxThresholdY–minThresholdY)
    else
        dY<<(distY - minThresholdY)/(maxThresholdY–minThresholdY)
}
```

Algorithm 1

Other input mechanisms can be supported by providing suitable values for the *maxDistX, maxDistY, initValueX, initValueY, currentValueX, currentValueY maxThresholdPercentageX, minThresholdPercentageX, maxThresholdPercentageY, minThresholdPercentageY.* For example, for a stylus-based interface, *maxDistX and* *maxDistY* can be set to the resolution of the screen; or for a D-pad to a period of time. The mapping functions allow input mechanisms to be used interchangeably.

Addressing the resolution problem

There are two pertinent problems. The first occurs on devices that have a relatively high resolution, but a small operating range. For example, when using a tilt controlled device, the tilt angle is restricted within a narrow range as the screen is only ideally visible within a very acute range. This effectively lowers the resolution of the device and results in the transitions in the scroll speed and zoom level being less fluid or smooth. The solution is to use a smooth interpolation function (see Algorithm 2)

Algorithm 2 uses cosine interpolation to produce smooth transitions between data points. *dX1, dX2, dY1, dY2* are two interpolation points in the x and y dimensions. This data can be obtained by recording the previous and current *dY* and *dY* values in Algorithm 1. *xVal* and *yVal* define where to estimate the value of the interpolated line, it is 0 at the first data point and 1 at the second data point. A value of 0.5 would estimate a value between the two data points. The number of increments between 0 and 1 determine the number of data points to estimate. *dXcurrent* and *dYcurrent* are the interpolated values.

The second problem occurs when the resolution of the device is extremely limited (e.g. each button on a D-pad can only distinguish between two level levels, on or off). The solution is to map translations in the x and y dimension to the duration each button is depressed. Of course, with this solution the reactiveness of the interface is lost as it is not possible to instantaneously change the scroll speed or zoom level. To counter this effect, a more aggressive mapping of the scroll speed and zoom level is needed. One benefit of this approach is that it allows smooth scroll and zoom transitions as data can be sampled at high rates. The sampling rate is constrained by the resolution of the system clock on the device.

```
dXcurrent << 0
dYcurrent << 0
xValTemp << 0
yValTemp << 0

if( 0 <= xVal <= 1)
{
   xValTemp = ( 1 - cos(xVal*PI))/2;
   dXcurrent = dX1*(1- xValTemp) + dX2* xValTemp;
}

if( 0 <= yVal <= 1)
{
   yValTemp = ( 1 - cos(yVal*PI))/2;
   dYcurrent = dY1*(1- yValTemp) + dY2* yValTemp;
}
```

Algorithm 2

Supporting single and dual-handed interaction

A byproduct of providing support for multiple input mechanisms is that the *AutoZoom* and *ManualZoom* interfaces can be used with single or dual-handed interaction. Both techniques can be used single-handedly using input mechanisms such as a D-pad or tilt-control, or dual-handedly with a stylus.

The *AutoZoom* and *ManualZoom* interfaces can be extended for use with other single or dual-handed input mechanisms. The *AutoZoom* interface could be used with other physical-dial type wheels such as on the iPod or the smartPad proposed by Rekimoto [102] for use in mobile phones. Meanwhile, the *ManualZoom* interface could be used with joystick-type mechanisms.

Designing control widgets

Two separate control widgets where developed for the *AutoZoom* and *ManualZoom* interfaces (see Figure 4.2 and Figure 4.3). The control widgets were designed to: increase navigation efficiency (i.e. there is a one-to-one mapping between the drag distance and the length of the green arrows); minimize the use of screen real estate (i.e. the design is kept simple and the arrows are kept thin and transparent so as not to occlude the information space); provide minimal distraction (i.e. the control widget is overlaid on top of the information space, so that the user's focus of attention does not need to shift as is the case with conventional widgets such as scrollbars); appears at the user's point of focus in the information space (i.e. with the stylus-based interface the user can select where to begin dragging, for other interfaces its placed in the center of the screen where the users point of focus is likely to be); and indicates the target location when completing a navigation action (i.e. the yellow dot is used to indicate the animation point).

4.2.2 Limited device capabilities

Cockburn and Savage [22] note that system performance is a critical issue for smooth scrolling and zooming interfaces. These systems need to be rendered rapidly and smoothly to work effectively. Their SDAZ implementation relies on graphics hardware acceleration to achieve high frame rates. Although most desktop computers include powerful graphic cards, very few handheld devices are equipped with this hardware.

Smooth scrolling and zooming applications also tend to be memory hungry as the entire information space is often loaded into memory. This can be extremely problematic on platforms such as Symbian OS version 8.0, where third party applications are only allocated a maximum of 2MB of memory.

In order to attain high frame rates on mobile devices, a cross platform (Symbian, Windows Mobile and Palm) 2D gaming library called GapiDraw[1] is used. Using this highly optimized library, the search techniques described in this chapter are able to

[1] http://www.gapidraw.com/

run smoothly and responsively at 30fps without needing any graphics hardware acceleration.

To overcome the memory problem, data is only loaded when it is needed. The information space is segmented into smaller portions that can be loaded and assembled when needed. To save even more memory, the information space is segmented at multiple zoom levels. High resolution segments are swapped for lower resolution segments as the view is zoomed out.

Multiple thumbnails are created for each photograph. The maximum thumbnail size is set according to the resolution of the screen. Successive thumbnail levels are created by scaling the maximum thumbnail size by a power of two (e.g. 2^0 (100%) 2^1 (50%), 2^2 (25%), 2^3 (12.5%) and 2^4 (6.5%). Each successive thumbnail level uses four times less memory. Lower resolution thumbnails are loaded into memory when the view is zoomed out or when the scroll speed is too large for higher resolution's thumbnails to be read in quickly enough. However, the more thumbnails levels there are, the more swapping is required and the more storage space is needed.

To guarantee that an image can always loaded quickly enough; the read speed must always be faster than the scroll speed. The read speed is largely dependent on the type of storage used (memory card or main memory) and the file format (compressed or uncompressed). Reading images from a memory card is much slower than reading them from main memory. Compressed images (e.g. JPEG) can be read in faster than uncompressed images (e.g. TARGA) as the file sizes are much smaller (e.g. JPEG images are up to 10 times smaller than TARGA images). However, compressed images require more processing time to decompress the image. Furthermore, the compression ratio tends to be better for larger JPEG images than smaller ones. For small thumbnails the compression may not be sufficient to counter the overhead of decompressing the file. In other words, it may be faster to get into memory a thumbnail that comes from a large uncompressed file than reading and decompressing a smaller file. Of course, the exact figures are dependent on hardware capabilities and the types of memory cards that are used. Using our hardware (HP4100 series pocket pc with ExtremeIII SD Card) the best combination that kept the ratio of the total read time to the scale more consistent was to use the JPEG format for the first three levels and the TARGA format to store the smallest thumbnail size.

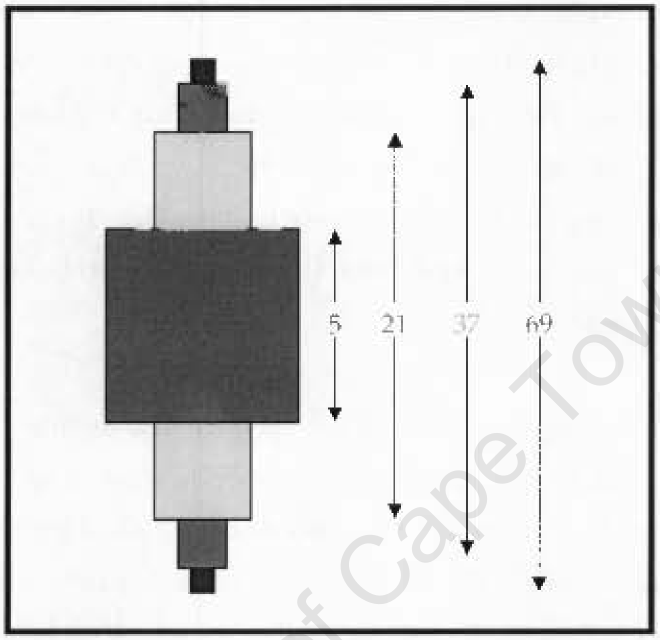


Figure 4.1: Memory buffer structure with four thumbnail levels (red = level 1 (100%), yellow = level 2 (50%) , green = level 3 (25%), blue = level 4 (12.5%).

The number of images at each thumbnail level is restricted to keep the total memory usage below 2MB. It is necessary to pre-buffer images at each thumbnail level so that they can be swapped in as the zoom level changes. Figure 4.1 illustrates the memory buffer structure. The four colors represent the four thumbnail levels (e.g. red represents the highest thumbnail resolutions and blue represents the lowest thumbnail resolution). To guarantee an image is always available, the lowest resolution thumbnails are read in first. Once the lowest resolution images are read in, the next level is read in, working its way to the highest resolution. If a lower resolution image is replaced by a higher resolution one, the lower resolution image is deleted immediately. Before reading in a higher resolution image, a check is performed to see if there are any lower resolution images that must be read in first. The highest priority is assigned to the lowest resolution thumbnails. A total of 69 images are pre-buffered, 5 at level 1 (100%), 16 at level 2 (50%), 16 at level 3 (25%) and 32 at level 4 (12.5%). The buffer is centered on the point of focus (i.e. the currently selected image).

Each frame is constructed using data from the buffers. This is done by determining which photographs need to be drawn in the frame and obtaining the necessary data from the buffers. The highest available thumbnail resolution is used. If the resolution is lower than what is required, the image data is scaled up to meet the requirements. However, this results in pixelization. This effect can be minimized by increasing the number of thumbnail levels. However, this introduces a greater swapping overhead.

The point of reading in multiple thumbnails is to guarantee that an image can always be read in quickly enough, even at the maximum scroll speed. Reading in only high resolution thumbnails is a waste of memory and processing power when lower resolution thumbnails are adequate (e.g. when the view is zoomed out). Having multiple thumbnail levels and buffers for each level helps us manage these resources. The buffering schemes are used to manage memory and ensure that only a few uncompressed large thumbnails are stored in memory at any given time. When the view is zoomed out, large thumbnails are down scaled in accordance to the memory buffer structure. For example, in Figure 4.1 the largest resolution thumbnail is down scaled to the appropriate resolution if it falls in levels 2, 3 or 4. The largest resolution thumbnail is kept if it falls in level 1 to avoid reloading it if the user decides to zoom in again.

4.2.3 Interface modifications

With the *AutoZoom* and *ManualZoom* interfaces, when the photographs are zoomed out, there is some white space on either side of the photo strip. A well known maxim attributed to Ben Shneiderman [9] states that "a pixel is a terrible thing to waste." The white space on either side of the strip provides an opportunity to place additional navigational aids. To maximize the use the limited screen real estate, we decided to modify the interfaces by adding folder information and temporal data (both of which are used extensively when searching for photographs [62] and are also readily available.). The folder information is gleaned from file structures stored on disc. It is displayed on a transparent information label shown at the top of the screen (see Figure 4.2). The folder information shows the path of the currently selected photo (i.e. photo directly beneath the yellow dot). Due to space constraints only a portion of the path is displayed, so the top-level folders are given priority. The temporal data is obtained by extracting EXIF data from each image header. Each photograph is time stamped by a digital camera at the moment of capture. The two horizontal bands on either side of the photo strip are used to show the date. The outer band shows the current year, while the inner band shows the current month (see Figure 4.2c). Changes in the year and month are depicted using alternating color bands (see Figure 4.2f). The information label at the top of screen also shows the corresponding year and month.

The scrollbars were removed to allow us to focus on understanding the problems with the *AutoZoom* and *ManualZoom* interfaces and to gauge their utility in sifting through large sets of images. In essence, removing the scrollbar, removes another factor for investigation. The scrollbar is replaced by a pointer. The two yellow arrows indicate the current position in the photo collection, where the top of the screen is the beginning of the collection (most recent photo) and the bottom of screen is the end of the collection (oldest photo).

Three changes were made to the *AutoZoom* and *ManualZoom* interfaces. The first change allows the photographs to be zoomed out to a level where a temporal overview can be seen on a single screen (see Figure 4.2c). The goal is to provide context and to investigate the utility of images when they are displayed really small. The second change was to remove the acceleration function from the *AutoZoom* technique. The motivation for this was to allow us to compare automatic verses manual control of the visual load. The third change was to lock the zoom level when reversing the scrolling direction. This was done to allow users to acquire targets more easily. Previously, when the scrolling direction was suddenly reversed, the rapid increase in the zoom level would cause a swelling effect that was very disorientating when acquiring a target that had just scrolled past.

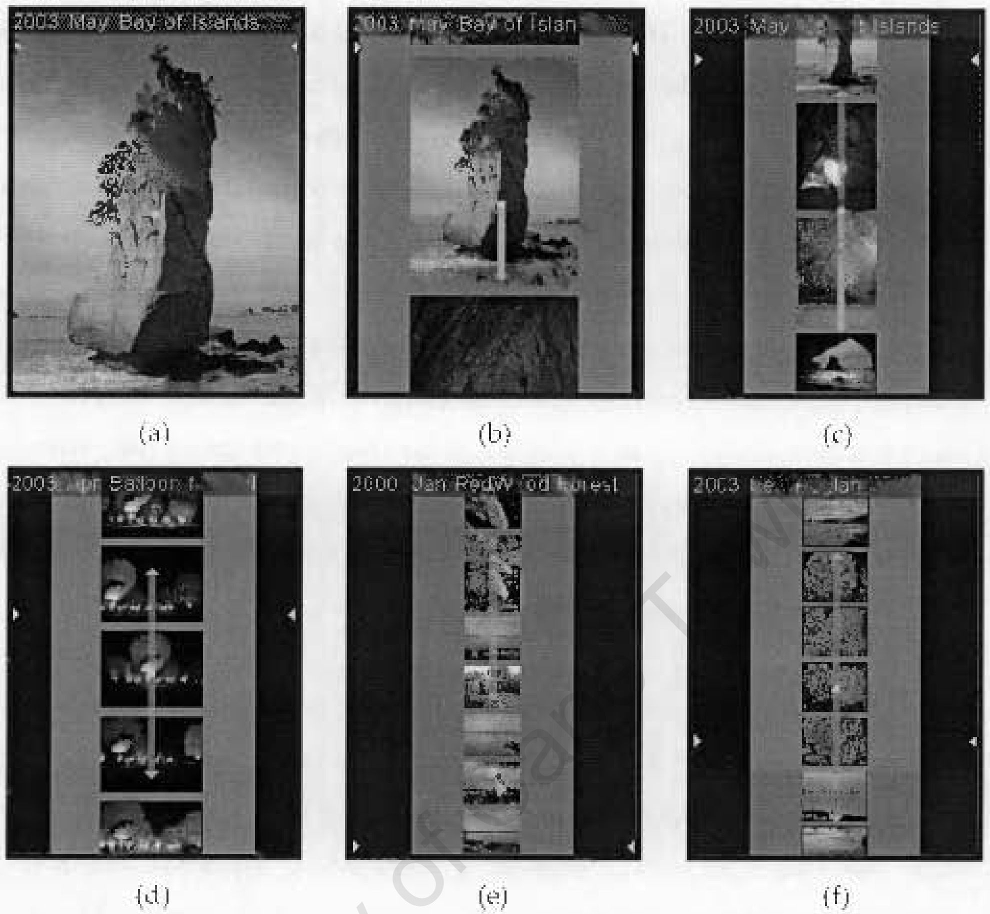
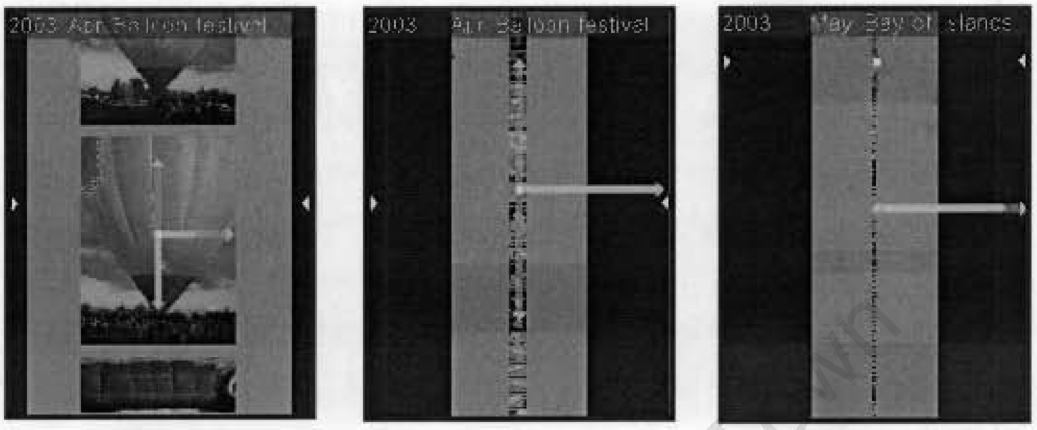


Figure 4.2: Screenshots of the augmented *AutoZoom* interface

The control widget was amended to provide feedback for the locking mechanism. When the scrolling direction is reversed, white arrows are shown (see Figure 4.2b and c). The white arrows indicate the distance a stylus has to be dragged before the zoom level is unlocked. The green bar shows the current drag distance. The zoom level is also locked on the *ManualZoom* interface (see Figure 4.3b and c). This is done to make it easier to scroll at specific zoom level.

For both interfaces the view is automatically animated in when the stylus is released from the screen. This provides another mechanism for adjusting the zoom level by simply lifting the stylus off the screen and then holding the stylus on the screen when a suitable zoom level is reached. The white arrows provide an indication of the current zoom level (see Figure 4.2 d, e and f). Besides these changes, the interaction is similar to the visual photo search techniques described in Chapter 3.



(a) (b) (c)

Figure 4.3: Screenshots of the augmented *ManualZoom* interface

4.3 Evaluation

The main aim of this study was to identify the limitations of the *AutoZoom* and *ManualZoom* interfaces in supporting the three search tasks (*events, singles* and *properties*) so that they can addressed to provide a better search experience.

4.3.1 Participants

Twelve participants (5 female, 7 male) took part in the experiment. Nine of the participants were students, while 3 were university lecturers. While having a larger and more varied sample would have increased the external validity of the results by making them more generalizable to the broader population, using 12 participants is an accepted compromise. Similar studies by Rodden and Wood [105] and Kirk et al. [62] have also used 12 participants. These papers are widely cited and have been accepted at CHI (the leading academic forum in the HCI community). Regardless of this fact, Lindgaard and Chattratichart [71] contend that the question of "how many users" is the wrong way to think about it. For example, in our study we are looking for a mismatch between our model of how people search for photographs and the users' mental model on the key and critical tasks. Framed this way, the key criterion that determines the number of problems that are found is not the number of users they are, but rather how many tasks the participants try. Lindgaard and Chattratichart [71] found that there was no reliable correlation between the number of participants tested and the number of usability problems uncovered. However, they found a significant correlation between the number of tasks that are evaluated and the number of problems that are uncovered. They conclude that with all other things being equal, the better predictor for the number of usability problems that are uncovered is the number of tasks that participants conduct in the study and not the number of participants that take part in the study. In our study great care is taken to create a broad and balanced range of tasks. In fact each participant conducts a total of 27 tasks.

To ensure the study was ecologically valid each participant was required to use their full personal photo collection, rather than using a stock collection or even a false subset that is edited to meet experimental criteria. The participants had varying numbers of photographs, with a mean of 1489.08 and a standard deviation of 1224.15. Having different sized photo collections in the study allows us gain a better understanding of the utility of the visual search photo search techniques, enabling us to obtain a broader view of the design space.

4.3.2 Method

There are various qualitative research methods. These include direct observation, introspection methods and query methods. To investigate the research goals, direct observation was used as it is good at identifying design problems. Direct observation involves observing and recording users interacting with the system.

The 'Think Aloud' protocol was also used to gain insights into decision processes (what participants believe is happening, what they are trying to do and why they decide on a particular course of action). As it is not natural to think aloud, participants were often encouraged to verbalize or explain their actions.

The validity and reliability of these methods (Direct observation with the 'Think Aloud' protocol) is dependent on how contrived or controlled the situation is. In this study, participants were required to conduct common searching tasks using their own personal photographs. The tasks were also tailored to each individual to ensure they were as relevant as possible.

Semi-structured interviews were used after the observation session to clarify observations and explore a small set of directed questions. They were used to gain information about unobservable phenomena such as thought processes. By having control over the line of questioning, we were able to examine any issue that was of interest. The interviews were very informal to give users the freedom to provide feedback on unexpected, yet relevant issues.

4.3.3 Materials

The *AutoZoom* and *ManualZoom* interfaces were deployed on a HP iPaq PocketPC 4100 series. A 1GB Extreme 3 SD card was used to store photographs. Two different input mechanisms were provided, stylus and D-pad.

Prior to the experiment, participants were requested to submit their entire photo collection. Each collection was processed by resizing the pictures to the same resolution as the screen. This was necessary to ensure that thousands of photographs could be stored on the 1GB SD card. Other than resizing each image, the photo collections were not altered in any other way.

An informal study with 12 participants was used to determine the number of thumbnail levels that would be appropriate for the HP4100 Pocket PC (which has a resolution of 320x240). The *ManualZoom* interface described in Chapter 3 was used for the study. Participants were required to look through their photographs and find the minimum level at which they were able to recognize events in their photo collection. The median scale value was 15. To avoid having to read in more data than is needed a minimum thumbnail size factor of 12.5 was chosen (in other words four thumbnail levels). The next step was to determine the maximum scroll speed at this zoom level. The *ManualZoom* interface was locked at a scale of 12.5. Users were then requested to find the maximum scroll speed at which they were able to recognize events. The median value was 407 pixels per second (or 9.25 images per second). This finding is consistent with research on human visual perception which shows that visual signals are summated over a period of approximately 120-125ms in daylight [22]. This means that users can just about process each image as it scrolls past. It is quite interesting that the maximum scroll rate at this zoom level is set fairly close the upper bound for visual tracking (i.e. the upper boundary where people are still able to track and identify images).

4.3.4 Procedure

The participants were randomly split into two groups, one for the *AutoZoom* interface, and the other for the *ManualZoom* interface. A conceptual model extraction [99] was conducted. The goal of the conceptual model extraction was to "extract how users interpret a completely new interface, given their existing mental models of how interfaces should work" [56]. Participants were asked questions such as, "what does this symbol represent", "how would you accomplish this task" and "how do you think this feature works." There are two benefits for using this technique. Firstly, the experimenter is able to obtain some feedback on the metaphors and ideas that work. Secondly, the participant is able to experiment with each feature in the application.

Following this, the participants were requested to scroll through their entire photo collection to familiarize themselves with the format of their photographs. They were then given 10 minutes to practice using the techniques. During this period they were encourage to think of photographs and look for them.

For the main experiment, the participants were required to complete 27 tasks. Being consistent with the study in Chapter 3, they were required to locate *events* and *singles* at short and long navigation distances. *Events* could also be small (3 or fewer photos) or large (more than 3 photos). *Singles* could have small or large features. The area of a small feature (e.g. a small child in a forest scene) was 1/8th (or less) of the total area of an image, while the area of a large feature was greater than 1/8th of the total area of an image (e.g. a skyscraper). In both categories we chose pictures that maximized the amount of space available, so the area for a small feature was as close to 1/8th as we could get and large features took up most of the space in a photograph. The tasks were read out by the experimenter. The participants were encouraged to think-aloud when performing tasks and were allowed to take breaks when needed.

Following the user observation session, an informal, semi-structured interview was conducted (see Appendix B for experimental materials).

4.3.5 Data captured

User actions were recorded for later analysis to avoid forgetting, missing or misinterpreting them. Pencil and paper was used to record events, comments and interpretations. A video camera was used to record all interaction on the device. The audio was also captured using a microphone. Both feeds were combined to generate a video for each participant. All the sections were video recorded and later transcribed.

4.4 Results

To ensure that the evidence collected and the interpretation was accurate and reliable, multiple verification strategies were used. According to Creswell [23] , the internal validity (the confidence in our explanation of experimental results) can be improved by various strategies such as: triangulation, deliberating rival hypothesis, including discrepant evidence, participant reviewing of the raw evidence, multiple researcher interpretation and analysis, and clarification of researcher bias. Several of these methods are used in our research. Multiple sources of information (i.e. direct observation, think-aloud method and interviews) were used to ensure the triangulation of data collection methods. Alterative explanations for the evidence were considered and checked with the actual users in follow-up interviews by showing them the raw evidence. Any discrepant evidence was also collected and assessed.

Section 4.4.1 begins by discussing how the participants went about locating *events*, *singles* and *properties*. Section 4.4.2 discusses how the interfaces were used to support these tasks, focusing on the folder information, temporal data, spatial cues and the two visual photo search techniques. Section 4.5 analyzes the results and discusses what is needed to support the search tasks more adequately.

4.4.1 Task types

Locating *Events*

We found that when locating events, landmark events and neighboring events were used as a navigational aid. When initially considering the position of an event, participants would think of the immediate surrounding events and would try to resolve their position relative to landmark events such as Christmas or birthdays. Landmark events were often be used as a course navigational aid, after which the participants would look out for neighboring events and use them to discern target locations displacement and direction. The position of a target event was continuously re-evaluated as more information was learnt and rediscovered.

When approaching a target location, participants would hover at what they called a "comfortable" zoom level. At this zoom level, they were able to get a sufficient overview to distinguish events and still pick up salient details in photographs. When more detail was required, they would zoom in momentarily to sample the event and then return to the comfort zoom level to continue searching for the event (assuming of course, that the target event was not found).

The process of finding photographs was similar for small and large events at short or long navigational distances. The process was rather more dependent on how well the event was remembered.

Locating *Singles*

When locating individual photographs, participants try to associate it with an event and then locate the event using the methods discussed above. Once the event has been located, the next step is to locate the target photograph.

The process of locating a photograph varies according to the type of features being searched for. Large distinct features can usually be picked out from the comfort zoom level. For less distinct features, participants would zoom in until they have sufficient data to identify the target image. If an image turns out to be a false positive, the view is zoomed out to obtain an overview.

When assessing many potential target photographs (e.g. photos containing the same object), participants would zoom in to gain as much detail as possible. They would then perform head-to-head comparisons until one was picked.

Locating *Properties*

Two different strategies were used to locate *properties*: a structured approach and a more sporadic approach. With the structured approach, the participants would systematically work through the entire photo collection from newest to oldest. Participants would move rapidly over irrelevant events and more slowly over relevant events. The information scent provided by the folder labels and thumbnails was used to guide users to "hot spots" (regions of the photo collections where likelihood of finding a target photograph was high). With the sporadic approach, participants worked through event folders from one likely event folder to the next. The decision of where to go next was not fixed, but evolved as more was discovered and learnt. Often visiting one hot spot would trigger memories that would form the basis for the next phase of exploration. Sometimes participants would navigate to a wrongly remembered event. In this case, they would backtrack and go down an alternative search path.

4.4.2 Contextual information

Folder labels and file structure

Event folder names were used to locate events in all photo searching task types. Participants preferred to use folder labels to distinguish events as opposed to scanning the photographs. We observed that event folder names contained different types of information such as, who attended the event, what they were doing, when the event occurred, where the event took place and why the event was so significant. Not all these categories (who, when, where, what and why) were always used. Participants usually choose the ones that were most important to them. The fact the users relied on this information shows that this metadata contains valuable information that is useful at search time.

We found that not all folders were event folders. Some folders contained special categories of pictures (e.g. scenic pictures) or events (e.g. weddings). Sometimes these categories were artifacts from previous searches. This allowed subsequent similar requests to be processed quickly.

The hierarchical folder structure was used for multiple purposes. It was used to organize events into main events and sub-events. Participants stated that this was done when importing photographs into the collection. Main events were given more generic names that served as an overview for sub-events. Sub-event names were given more descriptive names to distinguish them. In essence, the hierarchy was used as a level of detail filter, by providing an overview of the photo collection at the top of the hierarchy and allowing users to drill down the hierarchy to obtain more specific details. Participants would often scan through the top level folders to find an appropriate starting point. Selecting a main event effectively filtered off the rest of the photo collection. The hierarchical structure was also used as an archiving mechanism. Some participants would archive photographs at the end of each year, by placing them in a corresponding 'year' folder. This enabled them to have much easier access to the current photographs and also enabled them to search through older photographs by selecting the appropriate year.

Temporal information

The temporal arrangement was also useful for locating events. As these participants explain:

"Organizing photographs according to time is really useful, especially if you don't label your photographs. It's the only way you can think of finding them. My cameraphone doesn't allow me to organize my photographs, well probably because I can't store that many. But anyway, I'm reliant on the date when searching through my pictures." [Participant 4]

"I think of my photographs in terms of events and then try to remember when the event occurred. When I can remember it precisely, I go immediately to the date. When I can't remember it precisely, I try to narrow it down to a year or month. I also try to think of other events that happened around that time. " [Participant 3]

Events naturally occur sequentially in time. As a consequence, there is a unique temporal location for each event, assuming of course that events that can be broken down to sub-events are viewed as a single event. The temporal dimension provides an absolute dimension on which events can always be distinguished. For example, the same cannot be said of the location dimension as it is possible to make multiple visits to a single location.

Temporal navigation becomes more useful for larger and older collections where it becomes more difficult to recall event folder names accurately. For smaller collections, other cheaper (or less cognitively demanding) methods are used. As this participant explains,

"I think of my photos in terms of events, but don't really think of them in terms of time. I think this is because I don't have that many photographs. So normally, I just do a directory listing and scan through folders until I find what I'm looking for. This is far easier than searching by date as I don't have to remember when events occur in relation to one another." [Participant 2]

When temporal arrangements are applied naively they can break up categories of photographs or events. This is because categories will usually contain pictures from a variety of events, taken from different time periods. For highly categorical photo collections, an alphabetical ordering is usually more appropriate. As this participant explains,

"I have mixed feelings about the usefulness of organizing my pictures according to the date. I can see it is useful when you are thinking along the lines of when did I do that trip, but I'm a very categorical person, more categorical and less chronological. I think I would prefer an alphabetical listing to a date ordering and scan for the name. With an alphabetical listing

you can quickly skip to folder name because you know its relative position in the alphabet. I actually know my categories quite well. Once I find a category, the time information becomes more useful to find an event or picture" [Participant 11]

Temporal arrangements are also prone to synchronization problems that occur when using multiple capturing devices or when receiving photographs from other people. When multiple capturing devices are not synchronized in time, a temporal arrangement will often place photographs out of order. Having photographs out of order can be very confusing and disorientating.

Spatial information

The arrow pointers were used to passively portray the relative spatial location within the photo collection, allowing users to learn the spatial information if they attended to this cue. After interacting with the system, participants were able to remember the spatial positions of various landmarks events. When the spatial location was known and participants could not pinpoint an event temporally, we found that participants preferred to browse spatially rather than temporally as they felt it was more cognitively demanding to constantly resolve and reevaluate where you are temporally in relation to a target event. As this participant notes,

"I've got no idea when the Sombrero party is. I know it's at the bottom of the list. I'm scrolling to the bottom using the arrows as a guide. Normally, I would try to think of the date, but I know the pictures are near the bottom, so it's easier for me to just scroll to the bottom and then look for the pictures than trying to remember the date."

A few participants noted that they would have liked to use the arrow pointers like scrollbars to enable them to navigate rapidly to various locations in the photo collection. They also commented on the need for spatial cues that show the cost of navigation (for example, showing the number of pictures in folder or month).

4.4.3 Visual photo search techniques

The two visual photo search techniques were designed to control the visual blurring that occurs when scrolling rapidly. However, a few participants complained of eye fatigue and dizziness. This was largely due to the fact that participants were asked to perform 27 visual searching tasks consecutively, as part of the study. There were other smaller contributing factors though. One was the fact that the SDAZ technique tends to push for the upper bound for visual tracking to allow the fastest scroll rate possible. Perhaps a less aggressive mapping of the scroll-speed and zoom level would have been more appropriate. Another reason was the fact that participants did not control the visual load with the *ManualZoom* interface when navigating large distances. Participants were expected to zoom out to the minimum zoom level before scrolling. However, they scrolled as fast as they could at whatever zoom level they were at. Obviously, this led to visual blurring and could have contributed to the eye fatigue and dizziness. In fact, the ability to zoom was used to gain context, rather than a tool for controlling the visual flow at rapid scroll rates. When navigating rapidly to another location, participants did not mind the visual blurring. What was most important was the ability to get there quickly. Both techniques provide high granularity controls. This makes them ideal for moving small to medium distances, but not for navigating large distances.

The *AutoZoom* and *ManualZoom* techniques were most useful once an approximate location for the target was found. The *AutoZoom* technique was particularly effective when spurting between events. This was due to its ability to automatically control the visual load using very simple controls. However, the inability to zoom without scrolling made it difficult to "hover" over a set of photographs when inspecting them. As this participant explains,

"I like the AutoZoom technique simply because it requires very few movements. I can scroll rapidly through a set of photographs without paying much attention. It's particularly good for finding groups of pictures. But it's a little awkward to check out pictures because you can't zoom out without scrolling. This can be annoying when I want to get an overview, without moving away from a picture that I'm interested in." [Participant 1]

The independent controls provided the *ManualZoom* technique were more suitable for inspecting photographs. A few participants did however find them quite cognitively demanding. The participants that were more effective in using this technique would zoom first then scroll. As this participant explains,

"The ManualZoom technique is really good for checking out photographs. It's a nice comfortable action to adjust the speed and zoom level. It's difficult to get to grips with initially. I found once I separated scrolling and zooming it was actually quite easy. So I would first think to myself, what is the right zoom level. Once I found it, all I had to do was think about finding the right picture. I controlled the speed subconsciously, and kept my focus on the picture." [Participant 8]

Another benefit of these techniques is in supporting serendipitous browsing. As these participants explain,

"This technique [AutoZoom] is particularly good for finding photographs, but its even better for browsing though them. I like the way it forces you to see every single photograph. You get to see photographs you weren't looking for or had even forgotten about." [Participant 1]

"I really like the [ManualZoom] technique. There is something about it that attracts you to it. You just want to keep on playing with it. It's nice way of looking through my pictures. " [Participant 12]

They are also useful for storytelling, especially in situations where an individual is unsure about what other people are interested in and no special collection has been prepared in advance.

"These techniques are really good for showing people photographs. They show you everything. I think the list structure is well suited to interacting with someone, because they can follow the flow of pictures, the story. When you are chatting to people it is difficult to know what they want to see, and you have to wait for their cues. A slideshow is hopeless because it is difficult for other people to know what is coming and it takes a long time to get through pictures. With this new technique, you can now cover a lot of ground and people can pick up a point of interest that we can talk around." [Participant 7]

"Every so often you have to tailor the pictures you are showing depending on the circumstances. But your collection might not be tailored for a particular situation and you have to show photographs from all over your photo collection that you haven't categorized. These techniques provide a nice way of skipping to points of interest. You don't want duplicates of photographs, especially on storage limited devices, so instead you need a quick way to find the photographs that you want to show." [Participant 6]

"It's [ManualZoom] a lot better than my iPod photo. When I went for holidays in India I found it really useful to carry all my pictures with me on one device. It saved me from having to think about what photos to take. Because I had my iPod on me most of the time, I found that I showed my photographs quite often. Sometimes I knew that I would be seeing people that would ask me about certain things, so I made a special collection on my PC and copied it onto the iPod. With this interface [ManualZoom] I can get away without having to create a special collection. I can quickly skim through my photo collection, find a picture and show it, and then find the next one and show that. I think it would work really well when you bump into people and want to show them something, but haven't prepared the collection in advance. Of course I could try and do this on my iPod, but this interface does a better job of skimming through the pictures. With the iPod there is no easy way of adjusting the zoom level." [Participant 11]

One reason for the positive browsing experience is due the dynamic and interactive nature of these techniques and the fact that users are searching through content that they attach emotional value to. As this participant describes,

"There is something about the [AutoZoom] technique that makes it quite appealing. Its reactive, it does things, it animates, it's an engaging dynamic process that makes you want to play with it. I think the experience that people get is key, you are not only finding photos, you are doing something. It makes a normally tedious task much more interesting and fun." [Participant 3]

When reviewing photographs or making head-to-head comparisons a lot of small movements were required. As one participant explains, *"Sometimes when picking out a picture from a group of pictures, you need to jump backwards and forwards until you settle on one. Doing small movements to scroll up and down can be pain. It is difficult to skip through one to five photos. With an iPod you have more control, you just click five places and you are there. You have that tactile feedback and don't require great co-ordination skills. I think one way of improving this would be by providing a slideshow feature, where you can scroll from one picture to next and have the interface snap to a picture."* [Participant 8]

Participants also noted that there are times when they prefer to be actively engaged with the device or passively engaged with the device. When they are time pressured (e.g. when locating an image) they might opt for active engagement, whereas when they are less time-pressured (e.g. when browsing for serendipity) they might prefer passive engagement with the device. The reactive input mechanism (Stylus) was preferred for active engagement, whereas the non-reactive input mechanism (D-pad) was preferred for passive engagement with the device. With the D-pad it was possible to maintain a constant scroll speed without interacting with the device. The scroll speed is controlled by the duration a button is depressed. When a button is released the scroll speed is maintained constant until another button (any directional button or start key) is pressed. This feature allows users to find a comfortable scroll speed and then automatically scroll through the photographs without interacting with the device. This feature was mostly used to scroll through series of similar images.

Broadly speaking, participants did not find the photographs useful when they were minutely small. However, determining what the minimum cap should be is very difficult. For example, even when a picture is 2 or 3 pixels in width, colors can still be used to distinguish events with contrasting colors. Similarly, determining a limit for the maximum size is also difficult. For example, for very similar photographs a lot of detail might be required to select the better picture. What is essential is that users are able to select a suitable zoom level and are able to easily compare and contrast one or more photographs at any zoom level.

As a memory optimization, high quality thumbnails were replaced with lower quality thumbnails at high scroll speeds. All participants noticed that higher resolution thumbnails were swapped for lower resolution ones. Surprisingly, all participants stated that this did not interfere with their ability to search for *events*, *singles* or *properties*. At low scroll speeds high quality thumbnails were loaded so users could view as much detail as they wished. At high scroll speeds participants were not really looking at the pictures as they were moving too fast to make out what they were and also primary goal was to move quickly to the target location.

4.5 Discussion

4.5.1 *Events*

Based on observation and video analysis, we found that when searching a photo collection the participants would primarily think of their photographs in terms of events. This seemed to occur irrespective of the task type. For example, when locating a *single* they would first try to associate it with an event, then locate the event and finally locate the target photograph. When locating *properties*, they would think of events that were likely to contain target photographs and then navigate from one event to the next. These observations are consistent with a study by Kirk et al. [62] which have found that a common search strategy is often to think of the relevant event. Together this evidence suggests that the three photo search tasks rely on the ability to locate events rapidly.

We found that a number of different search strategies were used to locate events. Often participants would automatically think of the date as secondary process to identifying a target event. When the date was precisely known, they would navigate directly to the date. Otherwise, they would narrow the search down to a year or month by thinking of when it happened relative to neighboring events or landmark events. The task of locating an event was made easier by the fact that people organize photographs by events. Participants noted that the metadata that is associated with an event encodes the kind of information that they are likely to remember when searching for an event. We observed that this metadata was used extensively when locating target events. Some participants noted that they preferred to use the folder labels to distinguish events rather than scanning through the photographs. We also observed that some participants would organize events according to special themed categories. When searching for an event in a category, the participants would first identify the event, then think of an appropriate category and finally navigate to the category. This particular search strategy was problematic with our initial search interface as the chronological ordering broke up the categories. On occasion, we also found that users would navigate spatially to locate events. This evidence shows that multiple search methods are used to locate events. Ideally, users should be able to choose the most appropriate methods for the search task. Given that multiple search methods were sometimes used together for a search task (e.g. temporal navigation was used to narrow the search down to a year or month, then folder labels were used to identify events, and finally the *AutoZoom* and *ManualZoom* techniques were used to locate target photographs within the event) designers should investigate how to integrate them so that users can flip between them when performing a task.

There were a number of factors that influenced the search strategy. We found that for smaller collections users would often do an exhaustive scan to locate a target event as this required little effort and was not too time consuming. Participants with larger and older collections would often try to narrow the search down to a particular time period to avoid having to perform an exhaustive search. The user's memory of an event was one of the more observable factors across participants that impacted search strategies. For example, when an event was known precisely, users were able to locate it directly by looking for particular event information. In contrast, when an event was unknown, users performed a more exhaustive search where they would often examine each potential target event.

Based on the findings above, it is evident that the *AutoZoom* and *ManualZoom* techniques are not sufficient for locating events rapidly. They need to be integrated with techniques that allow users to directly access events when information about the event is well-known. For example, we found that for events such as birthdays and weddings, participants knew the exact date. In this instance, participants felt that the best way to locate a target event was by using a calendar interface, where they could input the date information and navigate directly to the event. Ideally, a search interface should try to make use of any information that a user has at their disposable, especially in this case where date information is readily available for each image. The *AutoZoom* and *ManualZoom* techniques also need to integrated with other techniques that narrow the search space when events are less well-known (especially as users felt fatigued when visually search through large numbers of photographs).

4.5.2 *Singles*

As stated above we found that when locating individual photographs users would try to associate them with an event. Users would first locate the encompassing event and then find that target photograph. When searching through the photographs, we found that the chronological ordering was useful as it helped users discern the displacement and direction of target photographs. When picking out photographs, users would often adjust the level of detail to acquire as much information as they needed. The *ManualZoom* technique was particularly effective in this regard as it allowed users to smoothly adjust the zoom level. The *AutoZoom* technique was less effective for inspecting photographs as it was not possible to zoom in or out without scrolling.

4.5.3 *Properties*

When performing *property* tasks we found that users would jump from one potential target event to the next using a structured or a sporadic approach. With the structured approach (where users systematically work through the entire photo collection), the jump from one event to the next is usually quite small. In fact, the *AutoZoom* technique was particularly well suited for this task. With the sporadic approach (where users skip from one likely event folder to the next), the distance between events is much more variable. To support this approach, navigation controls must allow users to adjust the granularity in relation to the navigation distance – for example, the iPod uses an exponential scale for scrolling through images. When rotating the wheel rapidly, the scroll speed increases exponentially allowing users to cover large distances quickly. The same can be done when an acceleration function is used with the *AutoZoom* technique (see Chapter 3). An alternative strategy for minimizing the amount of navigation is to group events together based on semantic properties.

To perform head-to-head comparisons with photographs from all over the photo collection, the entire process of selecting, reviewing, filtering and grouping photographs should be supported. The interface should also support navigation between these groups and the original event folders. For each selected photograph users should be able to flip back to the original event to get some context. The search history should also be maintained to show users where they have been or what areas need to be visited.

4.6 Summary

Section 4.1 outlines the research goals and motivates the need for a qualitative research methodology. Section 4.2 describes the prototype implementation. Section 4.3 outlines the experimental setup. Section 4.4 presents the results of a qualitative study. Section 4.5 analyses the results and provides design insights into viable solutions. This section provides some motivation for the research that is conducted in the next chapter.

Before developing another search tool as part of the next phase in the iterative design process, we felt it was necessary to clarify three issues that could have a bearing on the design. Firstly, we wanted to collect further evidence that users naturally organize photographs by events and provide descriptions for each event. If this is the case, then this would make way for a query-based search that takes advantage of this more light weight form of annotation. Secondly, we found that some participants grouped photographs according to special themed categories. We felt it was necessary to find out how common this practice is and also whether people predominantly organize by category rather than event. If this is the case, then this would suggest that focusing on events is too narrow as an organizing principle and that the schema needs to be broadened. Thirdly, we observed that not all event folders included the date or a description as some previous studies suggest [62]. In fact, the descriptions seemed to encode different types of information. If this is the case, then it would provide further evidence for the need to support multiple search techniques. The next chapter clarifies each of these issues.

Chapter 5

Photo organization and annotation practices

5.1 Introduction

In the previous study we found that that the ability to locate events is key to performing common searching tasks: locating *singles* involves first locating the encapsulating event; likewise, locating *properties* involves locating the most likely events and going through each of these events to locate matching photographs. We also found that the search strategy would differ depending on what was known about an event. The folder information was often used to locate events. Participants stated that this was because folder labels encode vital event information that is useful at search time and also because folders are used to delineate events. However, for a few participants we found evidence that suggests that not all folders encode event information. Some folders contain special categories of images such as "scenic pictures" or events such as "weekend trips." A question that arises then is: do people naturally organize photographs by events or by special themed categories. If the former is true, then we can support query-based searching by using the annotations for each event. If the latter is true, then it suggests that focusing on events is too narrow as an organizing principle and that the schema needs to be broadened to include other organizational schemes. In Chapter 4, we also noticed that participants would encode different types of information in the labels. However, we did not quantify this information. If indeed different types of information are encoded, it would provide further evidence for the need to provide multiple search methods that allow users to specify any information that is available to them. The primary goal of this chapter is to investigate each of these issues.

5.1.1 Current photo organization and annotation practices

Recently, studies on camera-phone annotation have found that user preferences for annotation are limited to a few favored images and some key information for each photograph (such as which event they belong to) [131]. Having to annotate each image can be time consuming and tedious as batches of images often depict the same thing [2]. Annotating immediately after capture can also be problematic. Firstly, it raises questions as to whether it is a socially acceptable practice [111]. The interruptions caused by taking a picture are often acceptable during social interaction. However, it is questionable whether the process of annotating after capturing can be seen in the same light. Preliminary studies have found that the lack of control over the timing of the annotation can be frustrating and stressful, resulting in negative feelings towards annotation [131]. Secondly, there are situations when it can be a hindrance or a burden, such as when multiple shots are required in quick succession to catch a passing moment. A better approach is likely to be one that considers users needs more closely. For example, by providing an easy way of annotating groups of photographs and providing a more laid-back approach to annotation, so that users can delay the annotation process to when they have 'time to kill'.

One such method for annotating groups of photographs on mobile devices involves adding a desktop component to the mobile application to make it easier to annotate and structure photo collections, when users are willing to do so. The desktop application is used to download photographs from the device and to organize them into one or more groups. Instead of annotating each photograph, key information is added to each group. Apple's iPhoto software[1] is a good example of specialist software that shifts the complexity of organizing photographs from the mobile device to the desktop computer. It allows users to store a copy of their entire digital photo archive on a device such as the iPhone[2] or the iPod Photo[3]. Any new images that are captured, or copied onto the device, are immediately available within iPhoto when the device is connected to a desktop computer (or laptop). Users can then organize these photographs within iPhoto and then apply the changes onto the device.

[1] http://www.apple.com/ilife/iphoto/

[2] http://www.apple.com/iphone/

[3] http://www.apple.com/ipod/ipod.html

This model for annotation has a number of benefits. It allows users to annotate images that are captured with a mobile device in a same way as their digital photographs. This enables users to merge their digital photo archives seamlessly with their camera-phone images. Using a familiar annotation scheme takes advantage of pre-existing mental models and user expertise in navigating through the photo collection. The descriptions provided for each group of photographs allow searching interfaces to be developed. In fact, a study of photo annotation on a cameraphone by Wilhelm et al. [131] recently found there is a 'need to develop hybrid solutions that integrate desktop and mobile application components into more complete and appropriate solutions.' They recommend adding desktop-based annotation capabilities.

Rodden and Wood [105] have looked at how people organize their print and digital photographs. Print photographs are often kept in their original packets. When users get round to organizing them, they usually place the most appealing prints into an album. Non-appealing prints are either thrown away or kept in their original packets. Albums are usually classified according to events, with one album per event. Within the album, photographs are placed in chronological order, with only a few changes in the order for aesthetic effect. Occasionally, users will write some important information on the back of the picture, such as the names of people in the photographs. Similarly, digital photographs are organized into event folders. Each event folder is given an appropriate name, usually consisting of date of capture and an event related description. Rodden and Wood [105] found that individual photographs are rarely annotated. The few images that do are usually sent to other people or posted on the web. It makes sense for users to group photographs according to events because users naturally associate photographs with events when searching [62]. A by-product of grouping by event is that photographs are also grouped according to time frames and common relevant locations [62].

All these studies provide motivation for a more lightweight model of annotation where a few meaningful keywords are attached to groups of photographs. As Rodden and Wood [105] point out, one mechanism for annotating photographs in this way involves using folders to group photographs according to events and folder labels to encode information about the event. Although these annotation practices have been observed before, no previous research that we could find has reported the fact that some people organize photographs into special themed categories. The goal of this chapter is to clarify whether or not people predominantly organize photographs into special themed categories as opposed to organizing them by events. In contrast to previous studies, it also assesses the type of information that is used to annotate groups of photographs. The findings provide corroboratory evidence to support some of the observations in Chapter 4 and assist us in formulating a set of more empirically grounded requirements that are feed into the next iteration in the design process which is presented in Chapter 6.

5.1.2 Contributions

There are two contributions:

- *Annotation practices for personal digital photographs.* Based on observations, semi-formal interviews and a thorough inspection of each participants photo collection we found that participants would sort their photographs by events when importing them into the photo collection. Photographs from other people that co-experienced the event were merged with the users own photographs from the event. We found evidence showing that some people organize photographs into special themed categories. However, even these participants would organize by events and would then place each event into an appropriate category. We found no evidence showing that pictures are sorted directly into categories. Although, we found some non-event related categories, such as "scenic pictures", these were usually created as part of a search task where the pictures were obtained from events folders. There were significantly more event groups than special themed categories.

- *A review of the type of information that is used to annotate events in a personal digital photo collection*. The annotations were analyzed according to five types of information: *who* is in the pictures; *where* the pictures were taken; *when* the pictures were taken; *what* the pictures are about; and *why* the pictures are significant. Each category (*who, when, where, what* and *why*) was broken down further into groups to conduct a detailed analysis. In our dataset, we found that on average 4.3 (s.d. 0.9) different types of information (*who*, *when*, *where*, *why* and *what*) were encoded in each photo collection. We found that the *who*, *where*, *when*, and *what* categories were used significantly more than the *why* category. The *where* category was used significantly more than the *who* category. There were no other significant differences between the different types of information.

5.1.3 Outline

Section 5.2 presents the experimental setup. Section 5.3 describes the results in terms of main objectives. Section 5.4 discusses some implications for future annotation and searching tools. Section 5.5 summarizes the major findings and provides an outlook for the next chapter.

5.2 Experiment

The key objectives of this experiment were:

- *To investigate whether people predominately organize their personal photographs by events or by special themed categories.* Studies by Rodden and Wood [105] and Kirk et al. [62] have found that people predominately organize photographs by event. In the previous chapter, we found that while most participants organized photographs by events, a few participants organized some of their photographs in to special themed categories. However, participants noted that categories were often the result of previous searches. Based on these observations and the fact the people that create categories also organize photographs by events, we hypothesize that people predominantly organize by event and categories are merely artifacts of previous searches. The alternative hypothesis is that some participants automatically sort photographs into categories (rather than events) and use categories as the predominant organizational scheme.

- *To assess the type of information that is used to annotate groups of photographs in a personal digital photo collection.* In a previous study, we observed that participants would encode different types of information about an event. Kirk et al. [62] found that a popular naming scheme was the date, followed by a description of the event. Considering five types of information (who is in the pictures; where the pictures were taken; when the pictures were taken; what the pictures are about; and why the pictures are significant), we hypothesize that people encode multiple types of information with the temporal dimension being dominant.

5.2.1 Participants

Twelve participants took part in the experiment. Five participants were female and seven were male. All the participants were recruited through advertisements posted around the university. The participants were taken from a variety of technical and non-technical backgrounds and had an age range of 20-47. Two of the twelve participants used specialist software to import photographs. The rest of the participants used Windows Explorer to manage their photo collection. Four participants used Photoshop to crop, resize, rotate, remove red-eye, create collages, stitch images and apply artistic effects. The participants were given a lunch voucher in return for their voluntary participation in the experiment.

Although, the results presented in this chapter may not be generalizeable to the population at large due to the small and selective sample, they are indicative of practice among college students who use Windows Explorer to manage their photographs and are therefore useful for design within this population group. In fact, some statistics on the recent social networking phenomenon, Facebook, show why it is important to consider this population group. According to a study by Emergence Media, 50% of all US users are college students, 21% are high school students and 29% are undeclared [33]. Furthermore, most Facebook users are between the ages of 18-24 [33]. We focus on Windows users as Windows usage accounts for 87.2% of total operating system usage, while Mac only accounts for 3.9% (and Linux 3.3%) [134]. All the participants used folders as a mechanism for annotation. Only two of the participants used tools to help them visually sort photographs into folders. We realize that the mechanism for annotation is different for Mac users as they use tools such as iPhoto to manage their digital collection.

With iPhoto the notion of folders is replaced with albums. Photographs are downloaded and placed in a virtual film roll. The photographs can then be organized into appropriately named albums. Keywords can be assigned to an individual photograph or to a group of photographs. Although the mechanisms for annotation are different, both approaches support the lightweight model for annotation, where photographs are organized into groups and keywords can be added to describe the photographs. In fact, the marketing for the latest version of iPhoto '08 supports our findings in the previous chapter and emphasizes the importance of organizing photographs by events.

"You don't remember photos by time and date. You remember them by events, like "Karen's graduation" or "Nathan's first bike ride." And that's exactly how iPhoto '08 organizes your entire photo library. So it's more fun to browse, and much easier to find a specific photo." [5]

5.2.2 Method

Multiple research methods were used to investigate the research goals. The data was collected using a set of interview and observation sessions. Some quantitative analysis was also conducted to follow up various points of interest. The data was normalized to make meaningful comparisons between individuals. Where applicable, analysis of variance tests (ANOVA) were used. Before conducting any ANOVA's the assumptions of normality and homogeneity of variances were tested to ensure they were not violated. When the ANOVA assumptions were violated, a non-parametric test was used (Kruskal-Wallis test). While these tests are less powerful than the normal ANOVA, they do not rely on any very serious restrictive assumptions concerning the shape of the distributions. When any assumptions are significantly violated the Kruskal-Wallis test is likely to yield more accurate results than an ANOVA.

In order to investigate the second objective, each photo collection was assessed according to the five dimensions of context (who, when, where, what and why). Participants were consulted to ensure the annotations were categorized correctly. Each category was broken down further into a number of groups (see Section 5.2.5). The quantitative data was collated and analyzed using basic statistical tools.

5.2.3 Materials

For the interviews and observation sessions participants were requested to use the computer on which they store their photograph collection to demonstrate their photo annotation and structuring practices. Four participants were interviewed at home as their photo collections were stored on their desktop computers at home. Eight participants were interviewed at university as their photo collections were stored on their laptops, which were used both at home and at work. Each participant was asked to bring along their digital camera with some photographs that had not been downloaded yet. An MP3 Dictaphone was used to record all sessions. Some notes were also taken to record the key activities that users were engaged in and also any interesting user behavior.

5.2.4 Procedure

Prior to the study each participant was requested to submit a DVD of their entire photo collection. This was used to prepare questions for each participant and allow us to assess the composition of the metadata that is attached to groups of photographs. At the beginning of the study, participants were asked to import photographs from their digital camera into their photo collection. They were told that we were investigating the mechanisms for annotation (i.e. using folders verses using photo management tools) to minimize participant bias, when we were actually interested in the how they organize photographs and annotate them. Following this, a set of more focused questions were used to probe their annotation practices (see Appendix C for experimental materials). Where possible, the participants were encouraged to use examples from their photo collection to demonstrate their answers. Some questions were more specific to each individual to allow us to gain further insights into their organizational schemes. The interview style was very informal and participants were free to express their views.

5.2.5 Data captured

For each participant, the following data was extracted from the folder labels:

- *Who is in the picture*: self, partner, family, friend, colleague, celebrity and pet.

- *When were the pictures taken*: year (e.g. 2007), month (e.g. January), day (e.g. 14^{th}), date (e.g. 24 October 1964), weekend, morning (e.g. early morning jog), evening (e.g. cocktail evening), night (e.g. night out), and calendar holiday (e.g. Independence day).

- *Where were the pictures taken*: country, state/province, city, town, suburb, street, house, landscape (e.g. Table Mountain, Camps Bay Beach), fauna and flora (e.g. World of Birds, Kirstenbosch Botanical Garden), hotel, restaurants and bars, community services (e.g. Cape Town International Airport, SPCA, Groote Schuur Hospital), education (University of the Cape Town), places of interest (e.g. Cable way, Waterfront craft center, Fish Market, Cape Wine Cellars), sports venues (Western Province Rugby Club), museums and galleries.

- *What are the pictures about*: abstract (e.g. green), food (e.g. sushi), artistic (e.g. graffiti), botanic-nature-landscape (e.g. orchid, snow, waterfall), objects (man made objects such as car or clock), society (e.g. festival, travel), sports and games (e.g. basketball), technical (e.g. molecule, hardware), and zoology (e.g. giraffe, wildlife). The groups in this category are based on the categories in the CEA-CLIC image database [82], which are also used in the MUSCLE categories and keywords for image annotation [43]. We only use the categories that are applicable to our dataset.

- *Why are the pictures significant*: milestone (e.g. *21^{st}* birthday), unique occurrence (e.g. *first* trip overseas), and unusual (e.g. *total* eclipse).

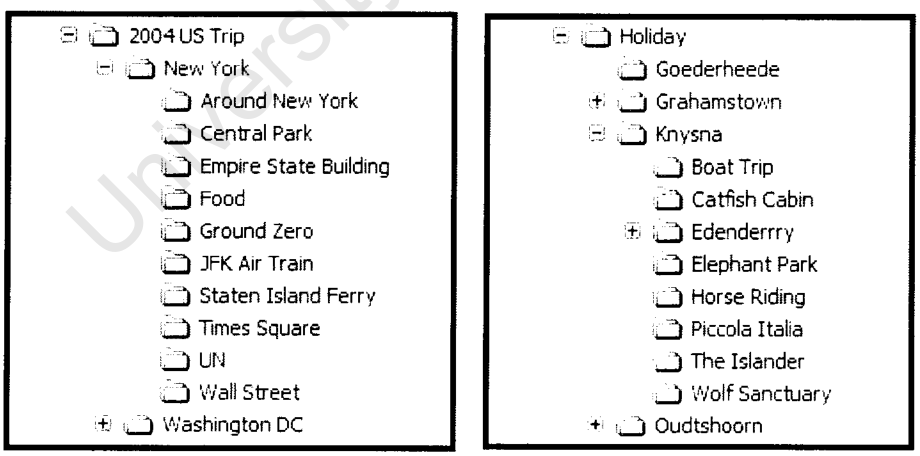
These categories are not exhaustive. It is possible to add more groups to each category. That said the groups shown above were adequate for the dataset that was used in this study.

5.3 Results

Section 5.3.1 presents our findings on how people organize their photographs. The data is based on observation and also on the semi-formal interviews. As we discussed above, the participants were encouraged to show examples to demonstrate their answers. Where possible, statistical data is also used to back up the observations. The data in the section directly relates to the first research question. Section 5.3.2 presents our finding on the type of metadata that is attached to groups of photographs. The data in the section relates to the second research question.

5.3.1 Photo organization

We found that when photographs are downloaded from the camera, they are usually placed in a temporary folder. The photographs are then visually sorted into event folders. Events are sometimes broken up into main events and sub-events as shown in Figure 5.1a. Sometimes, events are categorized according to a special theme (see Figure 5.1b). The photographs are organized by events and then placed into an appropriate category.



(a) (b)

Figure 5.1: (a) organizing photographs into main events and sub-events, (b) grouping similar types of events into a special themed category.

Not all themed categories contain events. Some categories contain a group of related images, such as pictures of pets. While event categories are usually created during the importing phase, picture categories are usually created as part of a search request. In the dataset, category folders account for 9.41% (s.d. 5.58) of all the folders a photo collection. The remaining folders are all event folders. The set of category folders (that make up the 9.41% of a photo collection) are comprised of two types of categories, picture categories and event categories. Picture categories account for 70%, while event categories account for 30%.

Event		Property		Arbitrary		Purpose-driven	
Trips	3	Partner	4	Miscellaneous	12	Temporary	3
Family	2	Location	4			Small size	2
Friends	2	Pet	3			For Gran	2
Work related	2	Friends	3			To Print	1
Birthday & parties	1	Scenic	2				
Scenic locations	1	Self	2				
Out and about	1	Neighborhood	1				
Weddings	1	Work related	1				
Balls	1	Cars	1				
Pets	1	Family	1				
		House sitting	1				
		Nature	1				

Table 5.1: Categories in the data set (the number indicates the frequency of each category)

Three different types of picture categories were found. The first type of picture category consists of handpicked photographs with a common theme or property, such as pictures of family or friends. The second type of picture category consists of arbitrary (or miscellaneous) photographs that are hard to classify in any of the existing categories. The third type of picture category contains a group of pictures that a have specific purpose or action attached to them (for example, "for grandma" or "to print"). In the dataset, 55% of the picture category is made up of the *property* category, 29% is made up of the *arbitrary* category, and 16% the *purpose-driven* category.

Table 5.1 shows the categories that were found in the data set. The categories are sub-divided according to event and picture categories (*property*, *arbitrary* and *purpose-driven*). Some of the original labels have been replaced with more generic names to make the categories more understandable. Popular event categories include: holidays; family get-togethers; social events (or celebratory functions) such as birthdays, engagement parties or weddings; and work related functions such as conferences or Christmas parties. Common *property* categories include: pictures of special people (or pets); favorite spot or location; and artistic or scenic pictures. The *arbitrary* category contains a single category called "Miscellaneous pictures" which contains pictures that are hard to classify. When an *arbitrary* category doesn't exist, the photographs are usually stored in the root directory. As this participant explains,

"Images that don't fit anywhere usually end up in the root directory. I accumulate them and shift them into folder containing similar ones. If they irritate me I might even delete them" [Participant 9]

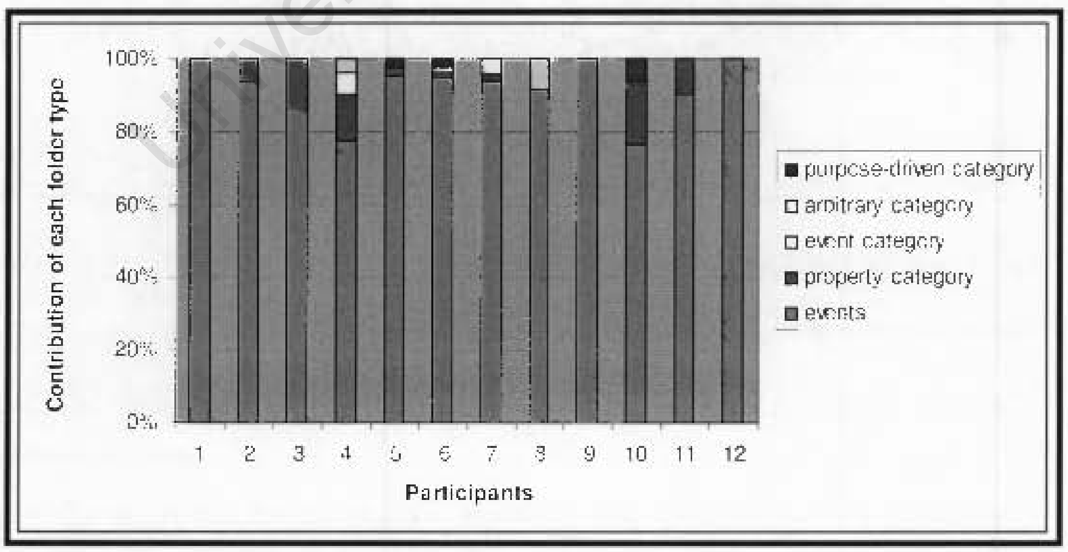


Figure 5.2: The distribution of folder types across participants

The *purpose-driven* categories will often have the users intention explicitly stated in the folder label (e.g. 'for gran', 'to print'). These categories are normally deleted once the goal has been accomplished. Occasionally, the pictures are kept because of their artistic value or because they contain a good summary of a particular topic or event.

Figure 5.2 shows the distribution of folder types for each participant. For each participant the bar graph shows the proportion of event folders and category folders (*purpose-driven*, *arbitrary*, *event*, and *property*). Each participant had significantly more event folders than *property* folders (Kruskal-Wallis, KW=15.36115 p=0.0002). They also had significantly more *property* categories than *arbitrary* categories (KW=4.613569 p=0.0317) and *purpose-driven* categories (KW=5.612311 p=0.0178). There were no other significant differences.

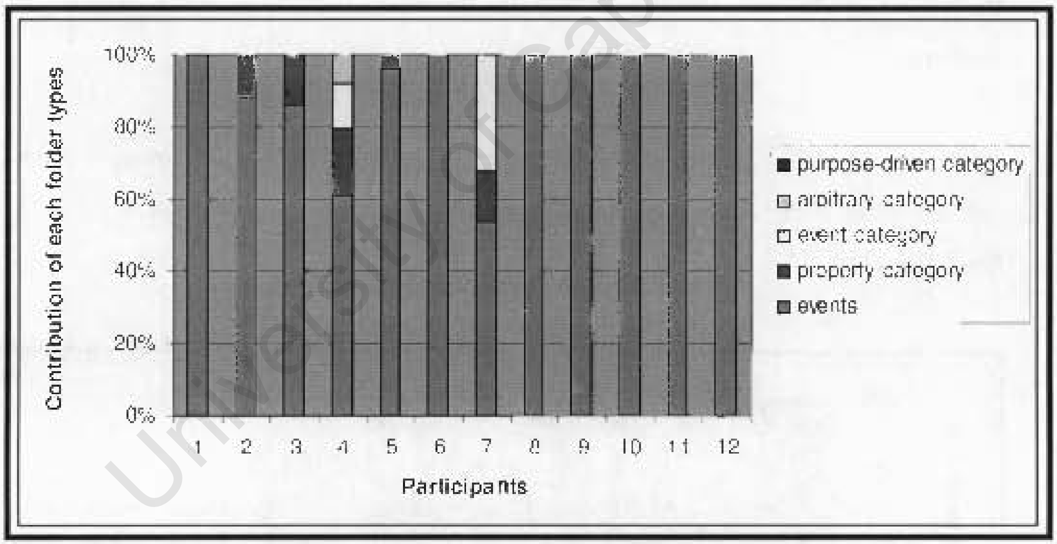


Figure 5.3: The distribution of folder types across participants. The figure only shows folders from the top level of the hierarchy.

Figure 5.3 also shows the distribution of folder types for each participant. In contrast to Figure 5.2, the bar graph only shows folders in the first level of the hierarchy as this is where you would expect to find more category folders. The bar graph clearly shows that some participants (see participants 4 and 7) do organize their photographs according to special categories folders. The important point to note here is that when importing photographs in the collection, any photographs that are captured by the participant are first sorted into event folders and then placed into an appropriate event category if one exists. According to participants, the role of the category in this case is to reduce the search space. They first think of the event and then the category that contains the event. Once the category has been located the event can easily be found by scanning through the reduced set of folder labels. Participants 4 and 7 note that it is far easier to remember category names as they relatively few in comparison to the number of event folders. As long as they have an idea of the event they can go to the appropriate category and search through the reduced set of folders for the event. It is also worth noting that participants 4 and 7 had the largest photo collections in our dataset.

All the users in our dataset accumulated photographs from other people such as friends, family or colleagues. In our dataset, 32% (s.d. 22.96) of all folders in a photo collection contain photographs from other people. This result is particularly interesting as it relates to the device synchronization problem that was observed in the previous chapter, where photographs from other people are placed in the wrong chronological order when cameras are not synchronized in terms of time and date. Although the actual figure (i.e. 32%) above might be higher than expected (due the sample comprising mainly of university students), the significance of this result is that it highlights the synchronization problem and the need for adequate solutions to address to problem, particularly for groups of people that share a large proportion of the photographs that they capture.

When participants import photographs from other people, a common strategy is to sort through the photographs for interesting pictures. Some participants are very particular about keeping their photographs separate and make a conscious effort to do so.

"If I get pictures from my friends, say of a birthday party we all went to. I'll create a main folder for my pictures and then sub-folders for everybody else's pictures. I'll sort through their pictures and keep the ones I want. I don't merge photographs because I like to see what photos I took." [Participant 2]

"When getting photos from other people I delete the common ones and only keep the unique or nice ones. I create a folder with the both the event and persons name. This folder is then placed under the main event. I keep photos separate because I like to know who took what." [Participant 12]

Other participants simply merge photographs from other people with their own photographs.

"When I get pictures from other people, I normally just merge them with my own photographs. It saves me the hassle of having to navigate through lots of folders to find them. I don't really feel the need to differentiate my photos. I select the photos that I want or like or are relevant to me and merge them with my photographs" [Participant 11]

In fact, some participants separate some events, but merge others.

"Sometimes I mix photos from other people with my own pictures. I usually merge pictures when they are quite similar. Other times I create a sub-folder with the person's name. I do this mainly when I want to see the photos each person took. Like when we go away for a weekend, some of my friends like taking scenic pictures and others prefer to take pictures with people in them. Each persons pictures is like their own account of the trip, so its nice to keep them separate." [Participant 8]

When participants receive photographs of events they didn't participate in, the photographs are more disposable. A few pictures might be kept for their sentimental or artistic value.

"Sometimes I get pictures emailed to me of something I wasn't directly involved in. ...these photographs do not make it into my collection. This is because I don't like to share other people's photographs. I feel there needs to be a long term interest. The few that I keep have some sentimental value." [Participant 7]

"When friends email me some random pictures, I don't normally keep them. Sometimes I might pick out one or two pictures." [Participant 9]

Only 4 out of 12 participants annotated individual photographs. They did so by renaming the filename. Individual photographs were annotated to remember unique (or peculiar) names or when sharing them with other people (via email or the web). For each participant, the number of individual photographs that were annotated was less than 1% of the total collection. The general consensus was that too much effort was required to annotate each photograph and that a much better way of distinguishing them was simply by looking at them.

5.3.2 Photo annotation

We found that folder naming strategies were quite divergent. For some participants it was imperative that events appear in the correct order. Figure 5.4 shows an example of two naming schemes that were used to ensure that events are placed in the correct order. Windows organizes folders alphabetically by default. These folder naming schemes ensure that when folders are organized alphabetically, they are also placed in the correct date order.

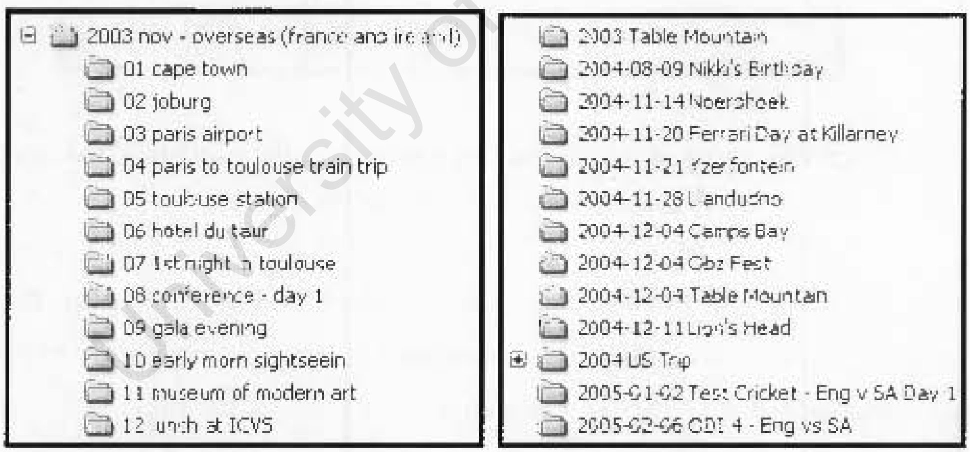


Figure 5.4: Encoding dates to ensure that events are arranged in correct date order

Other participants did not consciously use any naming scheme. The naming scheme was decided when labeling folders. The following comments suggest that this might be due to the photo collection size.

"I decide on a naming convention when importing. I'd like to have a better structure as its not ideal having to scan through a list of folders. But I get by fine. I guess I have a small collection, so it doesn't matter too much. Maybe when my collection gets bigger I'll have to think about how I name folders." [Participant 3]

"I don't have a scheme. I just do whatever I feel like on the day. I'd like to be more consistent, but my collection is small enough for me to know where things are."
[Participant 5]

Naming schemes	Mean (%)	Min (%)	Max (%)	Std.Dev.
name	74.23	5.88	100.00	26.16
name-year	5.51	0.00	33.33	9.47
year-month-name	5.00	0.00	60.00	17.32
number-name	4.59	0.00	34.12	10.40
name-number	4.37	0.00	27.27	7.71
year-month-day-name	2.70	0.00	32.43	9.36
name-year-number	1.36	0.00	16.36	4.72
day-month-year	0.76	0.00	9.09	2.62
year-month-day	0.73	0.00	4.88	1.72
year-name	0.45	0.00	5.41	1.56
year	0.22	0.00	2.63	0.76
month-day	0.08	0.00	0.97	0.28

Table 5.2: The likelihood of each naming scheme in a photo collection. The naming schemes were collected from the data set.

Table 5.2 shows the 12 naming schemes that were found in the data set. The most frequently used naming scheme is a description of the event, normally consisting of a few keywords. On average 74% of all folders in a photo collection use this scheme. The remaining 26% include a date component. Year and month date components are most likely to be found at the top of the hierarchy than at any other level in the hierarchy (using Kruskal-Wallis test at the $p = 0.05$ level). The year component is used significantly more than other date components at the top of the hierarchy (using Kruskal-Wallis test at the $p = 0.05$ level). For each user, the most frequently used naming scheme is used significantly more than other naming schemes at all levels of the hierarchy (using Kruskal-Wallis at the $p = 0.05$ level).

To get a better understanding of the type of metadata that is attached to groups of photographs we analyzed all the folder labels in our dataset in terms of five types of information: *who, when, where, what* and *why*. Figure 5.5 shows that 29% of the folders in our dataset specified *who* was in the picture, 53% specified *where* the pictures were taken, 24% specified *when* the pictures were taken, 39% specified *what* the pictures were about and 6% specified *why* the pictures were significant.

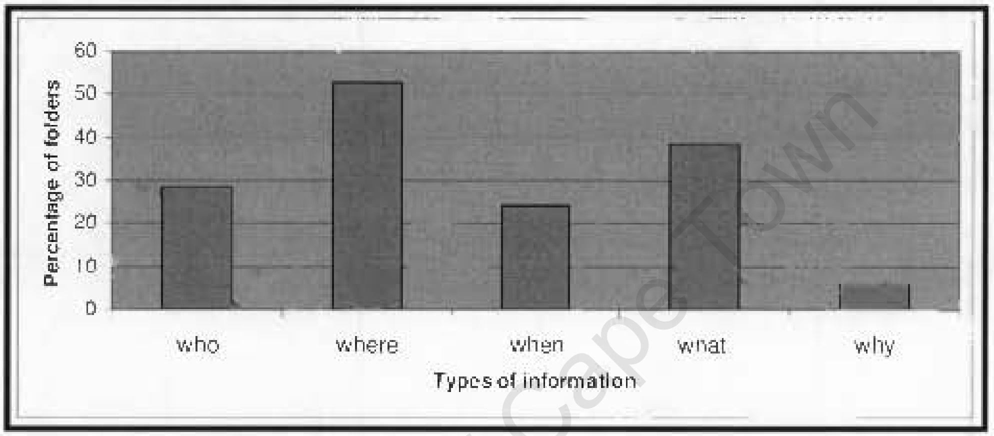


Figure 5.5: The different types of information (*who, where, when, what* and *why*) across the twelve participants in our study.

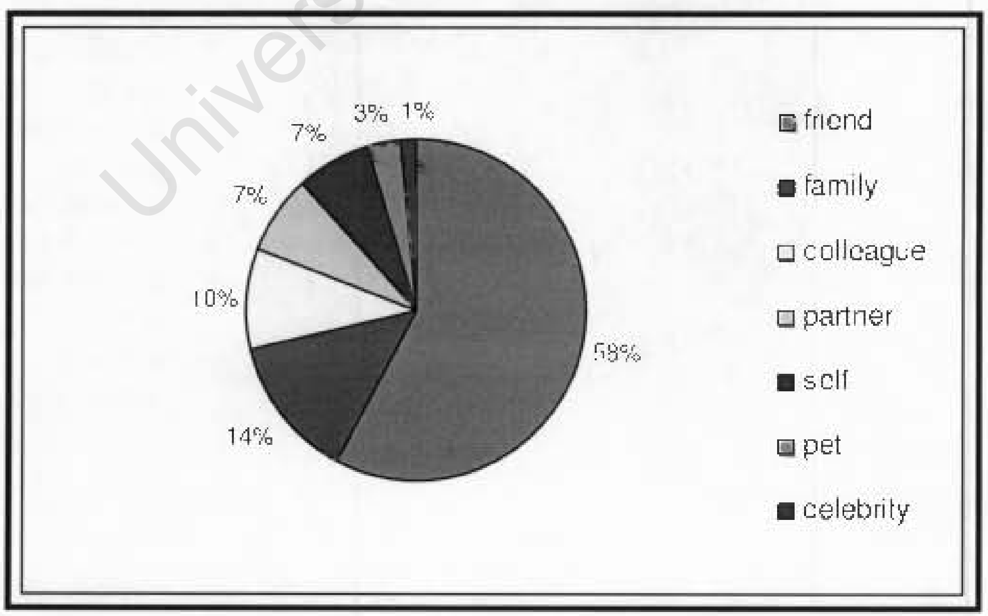


Figure 5.6: Breakdown of the *who* category

Figure 5.6 illustrates how the *who* category is broken down further into several groups. Folders are most likely to specify the names of friends (58%), followed by the names of family members (14%), colleagues (10%), partners (7%), self (7%), pets (3%) and celebrities (1%).

Similarly, the *where* category (see Figure 5.7) is broken down into the following groups: city/town (28.3%), places of interest (13.6%), landscape (11.4%), house (8.4%), restaurants/bar (6.5%), country (5.7%), fauna/flora (5.4%), education (4.3%), museum/gallery (4.3%), area/suburb (4.1%), community service (2.4%), monuments (1.9%), state/province (1.4%), hotel (1.4%), street (0.5%), and sports venue (0.3%).

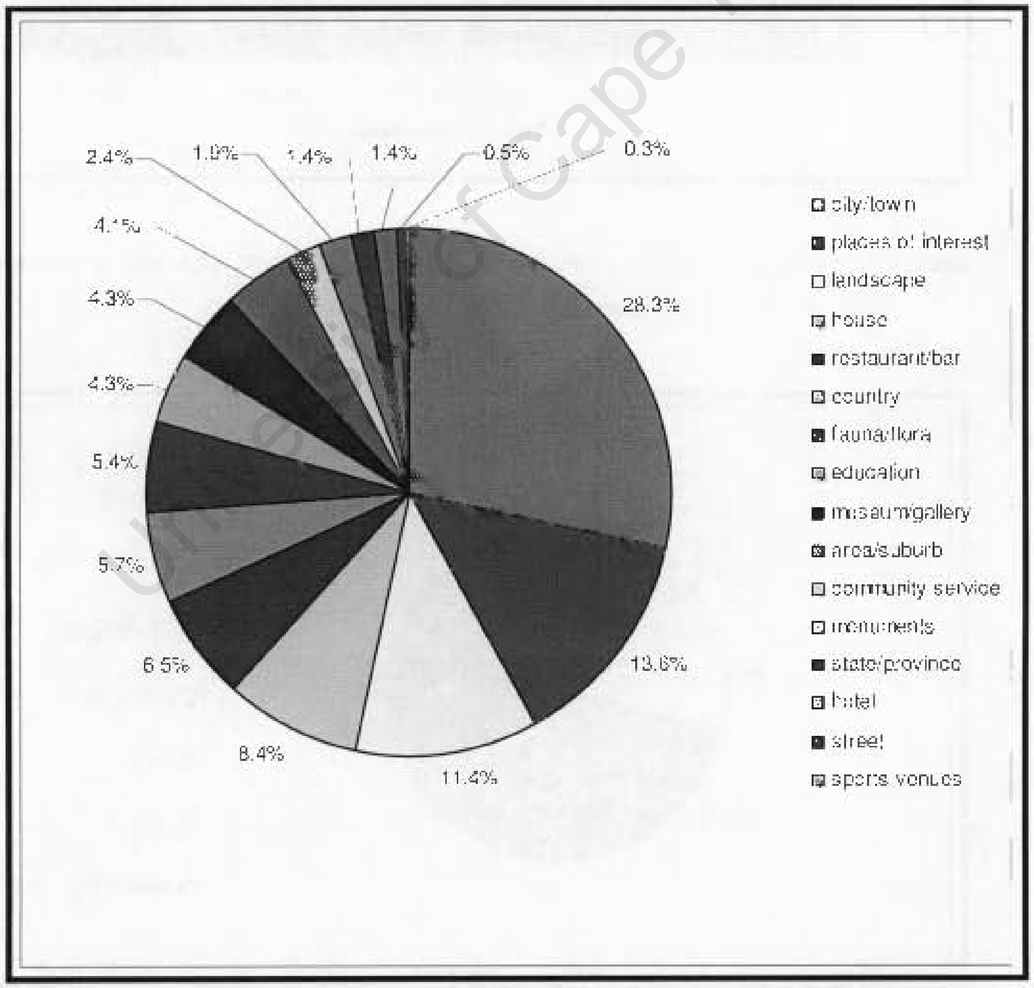


Figure 5.7: Breakdown of the *where* category

The *when* category is also delineated into a number of groups (see Figure 5.8). Folders are most likely to contain the year (49.8%) and month (26.4%), followed by the date (11.7%), calendar holiday (5.0%), night (3.8%), weekend (1.3%), morning (0.8%), evening (0.8%) and day (0.4%).

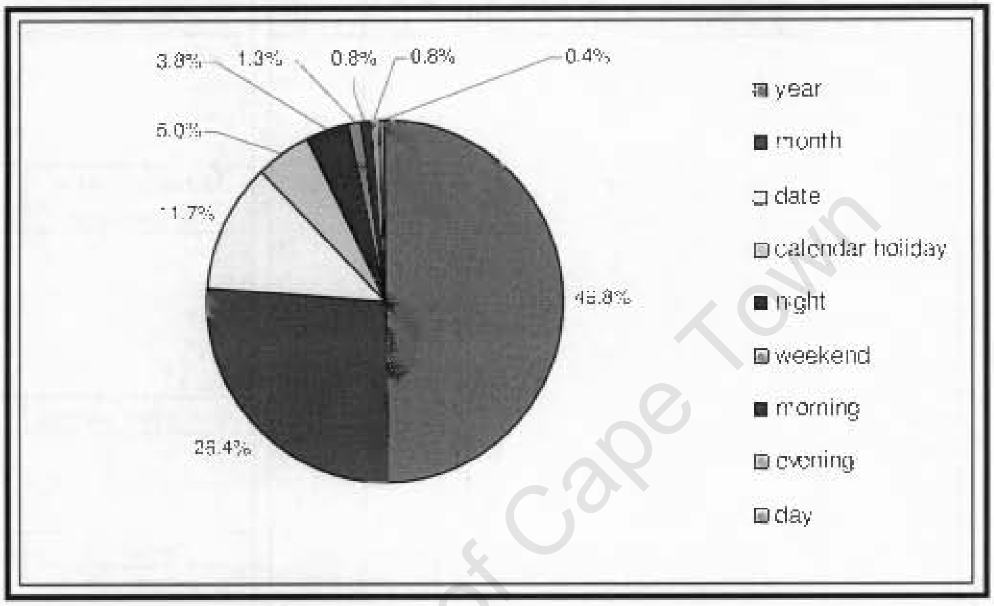


Figure 5.8: Breakdown of the *when* category

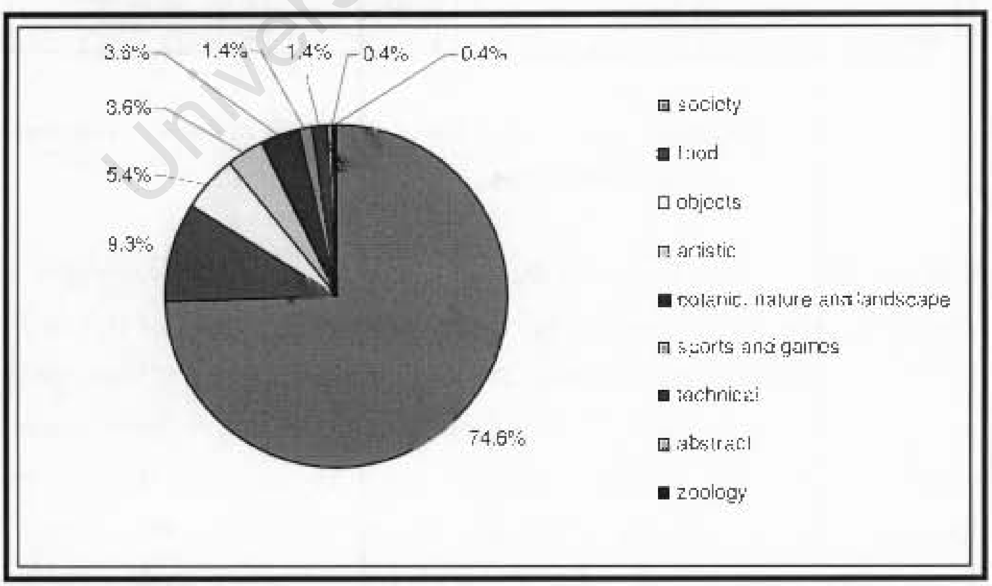


Figure 5.9: Breakdown of the *what* category

Figure 5.9 shows that almost three quarters of folder names in the *what* category fall in the *society* group (74.6%). The rest of the groups make up a quarter of the category: *food* (9.3%), *objects* (5.4%), *artistic* (3.6%), *botanic-nature-landscape* (3.6%), *sports and games* (1.4%), *technical* (1.4%), *abstract* (0.4%) and *zoology* (0.4%).

Society	F	U	Society	F	U	Food	F	U	Artistic	F	U
Birthday	36	10	Sight seeing	2	1	Food	7	4	Scenic	4	4
Trip	21	7	Supper	2	1	Fondue	4	3	Aerial	3	3
Party	20	5	Utopia	2	1	Sushi	4	2	Graffiti	2	1
Wedding	16	7	Anniversary	1	1	Cocktail	3	3	View	1	1
Conference	14	3	Breakfast	1	1	Cheese	2	2	**Botanic, Nature**		
Walk	12	5	Cleaning	1	1	Wine	2	1	**and Landscape**	**F**	**U**
Braai	9	3	Exam	1	1	Curry	1	1	Flowers	4	2
Farewell	9	4	Fashion Show	1	1	Edible	1	1	Snow	3	2
Graduation	7	6	Flying	1	1	Fish	1	1	Forest	1	1
Picnic	6	3	House sitting	1	1	Potjie	1	1	Garden	1	1
Dinner	4	4	Outing	1	1	**Objects**	**F**	**U**	Nature	1	1
Drive	4	2	Meeting	1	1	Fire	2	2	**Sports and Games**	**F**	**U**
House warming	4	3	People	2	2	Car	2	2	Cricket	2	1
Hike	3	2	Place	1	1	Phone	2	2	Horse Riding	1	1
Tour	3	1	Presentation	1	1	Train	2	2	Formula 1	1	1
Ball	2	2	Queer Party	1	1	Bricks	1	1	**Technical**	**F**	**U**
Dance	2	1	Speech	1	1	Objects	1	1	Equipment	2	2
Festival	2	2	Talent Show	1	1	Posters	1	1	Installation	1	1
Function	2	2	Team Building	1	1	Boat	1	1	Printing	1	1
Gala	2	1	Valentines	1	1	Displays	2	1	**Abstract**	**F**	**U**
Holiday	2	1				Ferry	1	1	Green	1	1
Lunch	2	2				Plane	1	1	**Zoology**	**F**	**U**
Open day	2	2							Dog	1	1

Table 5.3: Keywords in the *what* category. *F* is the frequency of use for a keyword. *U* is the number of users that use a keyword.

Table 5.3 shows all the keywords that were extracted for the *what* category. It also shows the frequency of each keyword in the dataset as well as the number of participants that used the keyword. The keywords are delineated into one of nine groups. The *society* group shows the range of activities that the participants were engaged in, with the dominant themes being celebration and travel. The other groups demonstrate the diversity of topics within a personal photo collection. Table 5.3 shows that there was some duplication of keywords between users. The fact that the frequency of use for a keyword is sometimes greater than the number of participants that use a keyword shows that keywords were are sometimes reused by participants.

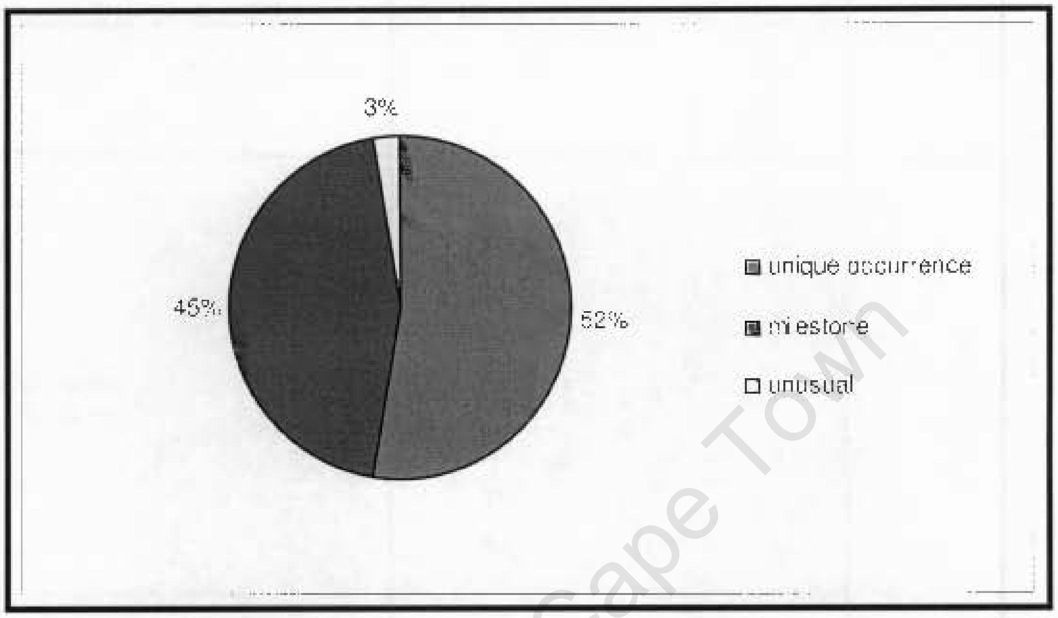


Figure 5.10: Breakdown of the *why* category

The *why* category (see Figure 5.10) consists of three groups: *unique occurrence*, *milestone* and *unusual*. In the dataset, the majority of this category is made up of the *unique occurrence* (52%) and *milestone* (45%) groups. The *unusual* group only accounts for 3% of this category.

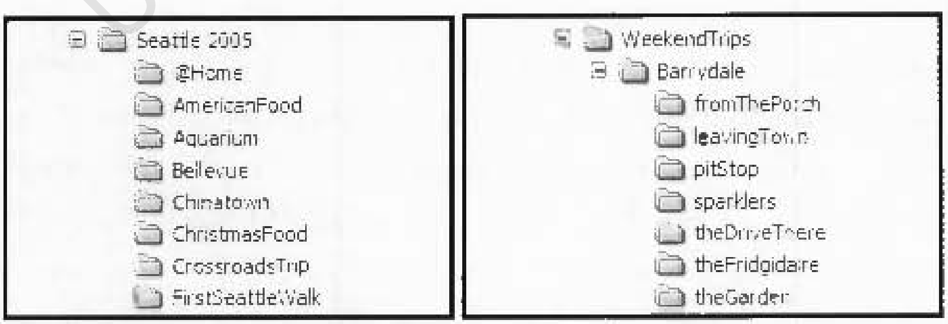


Figure 5.11: Two examples showing how information is encoded in a hierarchy

The hierarchy is often used to encode data about related groups of photographs by arranging them hierarchically into main folders and sub-folders (see Figure 5.11). Hierarchies allow key information about a group of related events to be added once. This avoids having to annotate each event with same information, reducing the effort required to annotate and avoiding duplication of information.

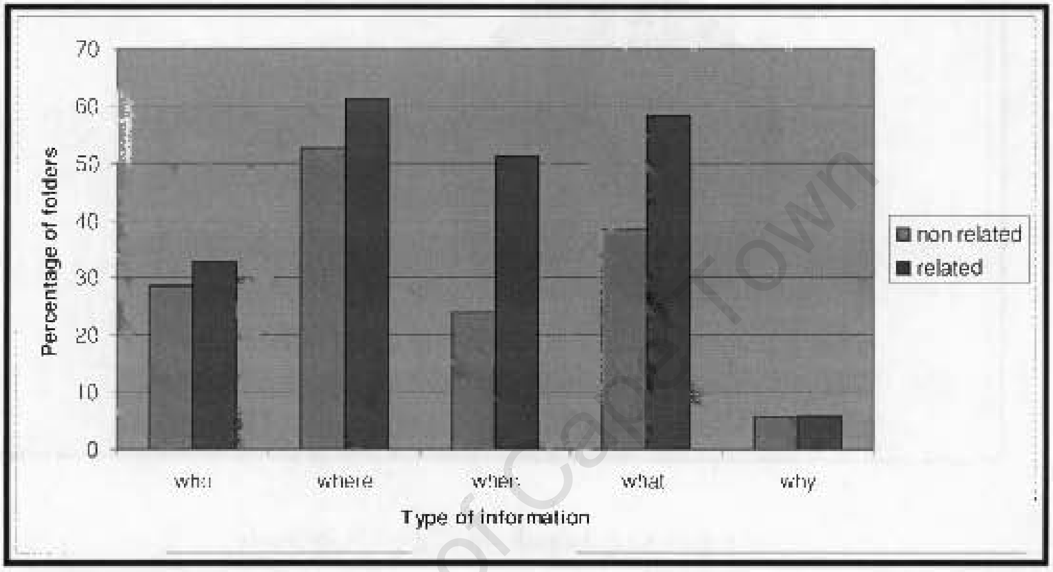


Figure 5.12: Encoding information in hierarchies. The graph illustrates the amount of information that is encoded using hierarchies for each type of information. The *non-related* group does not consider relationships between folders where as the *related* group does. The difference between the two is a measure of the amount of information that is encoded in a hierarchy.

Figure 5.12 illustrates the amount of the information that is encoded using hierarchies in our dataset. In the *non-related* group each folder is treated separately, where as in the *related* group, relationships with other folders in the hierarchy are considered. For example, let's consider the "Aquarium" folder in Figure 5.11. In the *non-related* group it treated separately, without considering other related information in the hierarchy (such as "Seattle 2005"). In the *related* group, metadata from the encompassing folder ("Seattle 2005") is also associated with the folder. For each type of information (*who, when, where, what* and *why*), the difference between the two groups (*non-related* and *related*) is a direct measure of the amount of information that is encoded in the hierarchy. The biggest differences are with the *when* (27.1%) and *what* (19.9%) categories, followed by the *where* (8.6%), *who* (4.3%) and *why* (0.1%). This represents an increase of 112% (*when*), 52% (*what*), 16% (*where*), 15 % (*who*) and 2% (*why*).

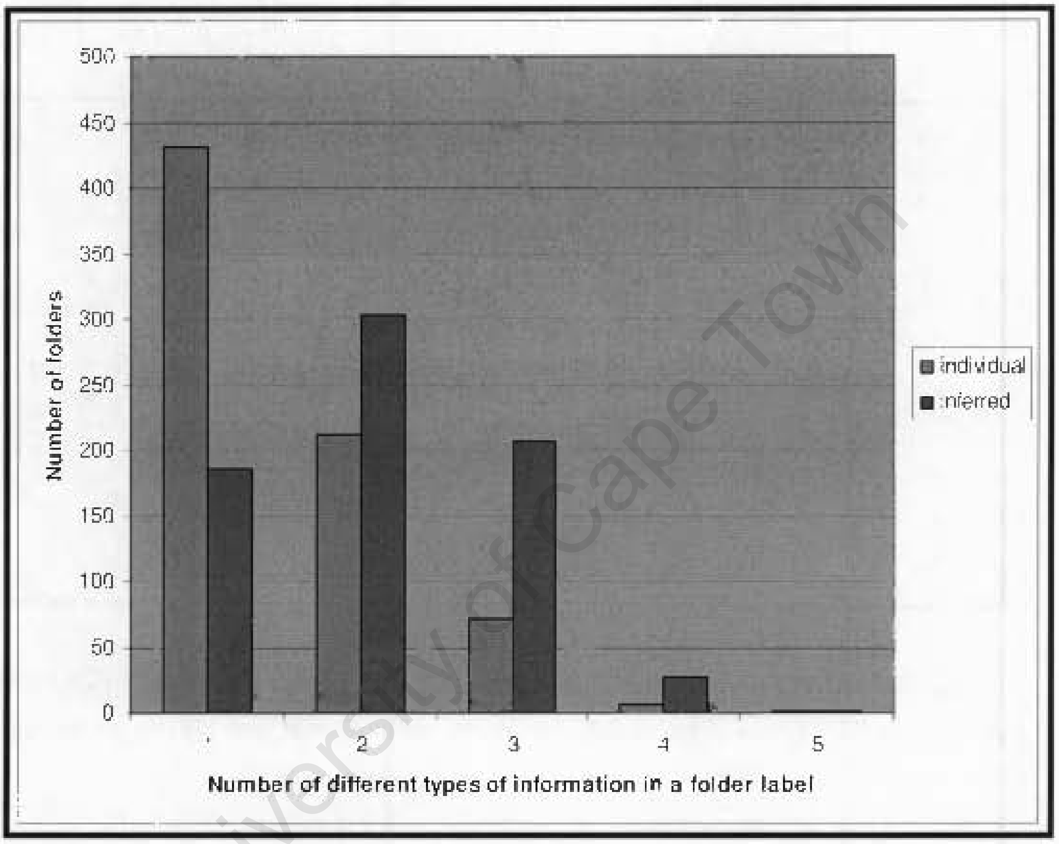


Figure 5.13: Encoding information in hierarchies. The graph illustrates the amount of information that is encoded using hierarchies by considering the number of different types of information in each label (*who, where, when, what* and *why*). The *individual* group does not consider relationships between folders where as the *inferred* group does.

Figure 5.13 provides another way of visualizing the amount of information that is encoded in a hierarchy. It does so by illustrating the amount of information that can be inferred when hierarchical relationships between folders are considered. In the *individual* group, each folder is treated separately, where as in the *inferred* group hierarchical relationships are considered. Within each of these groups, Figure 5.13 shows the distribution of folders based on the number of different types of information in each label (*who, where, when, what* and *why*). The decrease in count 1 and the subsequent increases for counts 2 (90 folders), 3 (135 folders), and 4 (20 folders) indicate the amount of information can be inferred by considering hierarchical relationships. In the *individual* group, 40% of all the folders in the dataset encode more than one type of information, whereas in the *inferred* group, 74% of all the folders in the dataset encode more than one type of information.

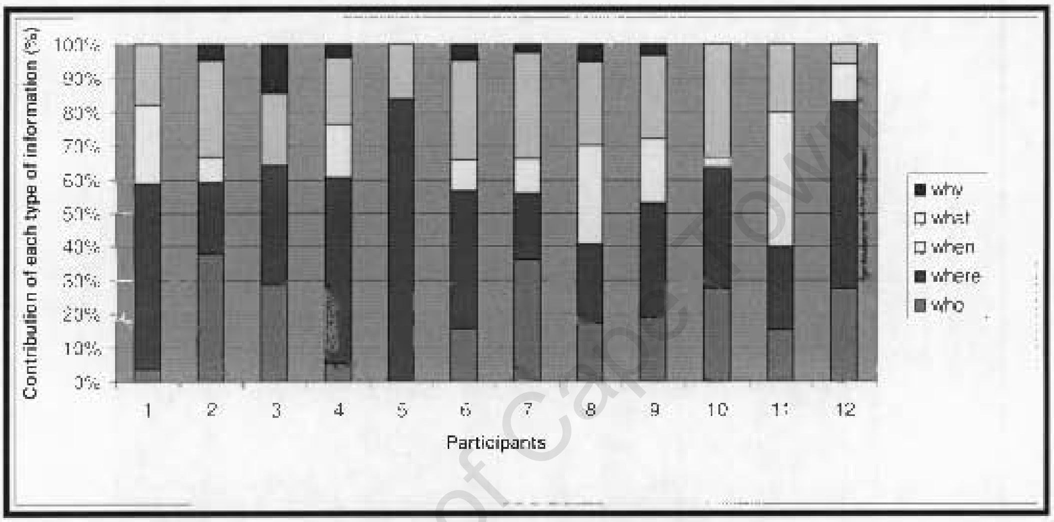


Figure 5.14: The different types of information that were encoded in folder labels by each participant. For each participant it also shows the proportion of each type of information.

Figure 5.14 shows the different types of information that were encoded in folder labels by each participant in our study. For each participant, it also shows the proportion of each type of information. On average, 4.3 (s.d. 0.9) different types of information were encoded in each photo collection. The *where* and *what* categories were present in each photo collection. Eleven participants specified *who* was in the photos, ten specified *where* the pictures were taken and seven specified *why* the pictures were significant. We found that the *who* (Kruskall-Wallis, KW=10.05827 p=0.0015), *where* (KW=17.43158 p=0.0001), *when* (KW=7.547089 p=0.0060) and *what* (KW=16.95073 p=0.0001) categories were used significantly more than the *why* category. The *where* category was used significantly more than the *who* category (KW=6.606578 p=0.0102). There were no other significant differences between the different dimensions of context.

5.4 Discussion

Considering the results in the light of the two hypotheses noted in Section 5.2

1. **People predominately organize their photo collection by event**. Based on observations, semi-formal interviews and a thorough inspection of each participants photo collection we found that participants would sort their photographs by events when importing them into the photo collection. Photographs from other people that co-experienced the event were merged with the users own photographs from the event. We found evidence showing that some people organize photographs into special themed categories. However, even these participants would organize by events and would then place each event into an appropriate category. We found no evidence showing that pictures are sorted directly into categories. Although, we found some non-event related categories, such as "scenic pictures", these were usually created as part of a search task where the pictures were obtained from events folders. There were significantly more event groups than special themed categories. The fact that all the participants organized photographs by events provides further evidence that supports an earlier observation in Chapter 4, where it was noted that people tend to think of their photographs in terms of events.

2. **People annotate events using multiple types of information**. In our dataset, we found that on average 4.3 (s.d. 0.9) different types of information. We found that the *who*, *where*, *when*, and *what* categories were used significantly more than the *why* category. The *where* category was used significantly more than the *who* category. There were no other significant differences between the different types of information. The fact that multiple types of information were used to annotate events provides further evidence for the need to provide multiple search techniques that allows users to specify any information that is available to them. In contrast with previous research [62], we found that the most popular naming scheme was not the date followed by a description. The most popular naming scheme was simply a description

of the event that was encoded using a few appropriate keywords. Furthermore, the temporal dimension was not the most dominant type of information.

5.4.1 Design implications for annotation tools

Hierarchical folder-based structures provide an easy way of delineating photographs into groups (using folders) and then encoding the relationships between groups (using a hierarchy). Users can attach key information to groups of photographs without having to replicate the information for each individual group. This reduces the amount of effort required to annotate and avoids duplicating information. However, there are some limitations in using folders and hierarchies to annotate groups of photographs. Firstly, it is not possible to assign an event to more than one category without having to duplicate information. Secondly it is difficult to add free text captions to a group of photographs, due to the limitations posed by folder naming conventions and the character length limit. Thirdly, there is no easy way of arranging event folders in the correct date order without explicitly encoding it in the naming scheme. Windows allows users to sort folders according to the modified date. However, the modified date changes when photographs within the folder are manipulated in some way. Although, photo management tools such as iPhoto provide this functionality, 10 out of 12 participants chose to use Windows Explorer because of its simplicity. Many photo management tools are packed with too many features and designed for multi-purpose use. Importing tools should not follow this 'Swiss army knife' approach [91] and should rather focus on doing one task well. They should take advantage of the fact that people are willing to group photographs and provide metadata that describes them.

In our dataset, we found that less than 1% of the photographs in a photo collection are annotated. Clearly, a major challenge for annotation tools is to motivate users to annotate their photographs. One way of doing so is to make photo annotation more fun and engaging. One approach is to turn photo annotation from a solitary activity in to a collaborative activity that encourages social interaction around photographs, and inadvertently get users to annotate photographs at the same time. For example, by posting photographs on social sites such as Flickr[1], so that different circles of friends can view them and comment on them. Some social sites such as Hi5[2] have added a competitive aspect to annotation by displaying popularity ratings, showing how often photographs are viewed and the number comments you have relative to other people. The goal is get users to 'spruce up' their profile by adding more pictures and witty comments to encourage other members to visit and post comments. Another more engaging approach gets users to annotate images by pitting them against each other in a game [125]. An alternative way of motivating users to annotate photographs is by making the benefits of annotation more apparent. Given that the primary reason for annotating photographs is to facilitate searching and ultimately sharing, importing tools should make the benefits of having a better structure and "richer" metadata more apparent. For example, by showing how annotation can enable compelling new ways for locating photographs. Importing tools should be interwoven with all the stages of the photo cycle (capturing, archiving, searching, and sharing). For example, in the study we found that event categories were normally created when importing photographs, where as picture categories were usually created as a result of a searching task.

5.4.2 Design implications for search tools

The fact that people organize photographs by event and annotate events with different types of information opens up new ways to support photo searching. The most obvious approach is one that takes advantage of the power of conventional information retrieval by indexing the metadata that is attached to each event so that users can retrieve events rapidly using a query-based search interface. Of course, query-based search techniques need to be integrated with other techniques that allow users to navigate through the photographs once an event has been found. Other techniques are also needed when the event information cannot be recalled or when the search requirements are more vague. Search interfaces should preserve the relationships between events such as main events and sub-events, and should take these relationships into account when indexing a photo collection.

[1] http://www.flickr.com/

[2] http://www.hi5.com/

5.5 Summary

Section 5.1 motivates the need to clarify some of the observations in Chapter 4. Section 5.2 describes the experimental setup. It begins by outlining the two research goals: to investigate whether people predominantly group photographs by events or by special themed categories and to investigate whether groups of photographs are annotated with different types of information (*who*, *when*, *where*, *why* and *what*). It also describes the methodology that was used to collect the data and the procedure that was followed for the experiment. Section 5.3 presents the results in terms of the two research goals. Section 5.4 reviews the major findings with reference to some observations in the previous chapter. It also discusses some implications for importing and searching tools. The rest of this section provides an outlook for the next chapter.

Two findings in this study were that people predominately organize photographs by events and often encode multiple dimensions of context about the event. This information can be used to support query-based approaches to complement the *AutoZoom* and *ManualZoom* techniques that allow users to visually search through the photo collection. This enables users to perform a query-based search when the keywords used to describe events are known and to browse for events when the metadata that was used to describe them cannot be recalled. In the next chapter we develop a search tool that integrates multiple search methods (*Keyword search*, *Timeline browser*, *Timeline filter*, *Hierarchical folder-based browser*, and the *AutoZoom* and *ManualZoom* techniques), based on the findings in Chapter 4. The design of each technique is focused on locating events. The next chapter brings together all the research in thesis by investigating how this tool is used to support the three searching tasks, taking in to consideration the fact that different search strategies that are used based on how much information is known about a target event.

Chapter 6

Integrating multiple search methods

6.1 Introduction

In Chapter 4 we found that when searching a photo collection the participants would primarily think of their photographs in terms of events. This seemed to occur irrespective of the task type. For example, when locating a *single* they would first try to associate it with an event and then locate the target photograph. When locating *properties*, they would think of events that were likely to contain target photographs and then navigate from one to the next. We observed that multiple search strategies were used to locate events. We found that the *AutoZoom* technique was particularly effective when spurting between events, due to its ability to automatically control the visual load using very simple controls. The independent controls provided by the *ManualZoom* technique were more suitable for inspecting photographs. However, we found that both techniques were inadequate for locating events rapidly and recommended that they should be integrated with techniques that allow users to directly access events when information about the event is well-known. For example, for events such as birthdays the date information could be used to locate the event. We also recommended that they need to be integrated with other techniques that narrow the search space when events are less well-known (especially as users felt fatigued when visually search through large numbers of photographs). In Chapter 5, we found that people organize photographs into event folders and provide metadata for each event. This metadata can used to create indexes for query-based systems that retrieve events rather than individual images, providing techniques to complement the *AutoZoom* and *ManualZoom* techniques. In this chapter, we develop a single search interface the integrates multiple search techniques (*Keyword search* for rapid access to events; *Timeline filter* to narrow the search space; *Timeline browser* to support searching by date; *Hierarchical folder-based browser* to delineate events and maintain relationships between events; and the

AutoZoom and *ManualZoom* techniques for sifting through photographs). The design of the search tool is focused on locating events. This is important as people naturally associate photographs with events when searching [62][105].

In this chapter, we assess the search tools ability to support the common search tasks. The study validates some of the observations in Chapter 4, enabling us to develop more empirically grounded guidelines for designing photo search interfaces.

6.1.1 Contributions

Three contributions include:

- *A single photo search interface that incorporates the best traits of a variety of tools to support search.* The search tool integrates four search methods (*Hierarchical folder-based browser, Timeline browser, Timeline filter, Keyword search, AutoZoom* technique and the *ManualZoom* technique). The design goal is to provide rapid access to events. The blend of different search techniques is designed to support users in locating events when they are known precisely (e.g. using the *Keyword search*) and also when they are less well-known (e.g. using the *Timeline filter* to narrow the search space). The search techniques can be used together to perform a searching task. Any constraint that is applied on one search technique is immediately reflected on the other techniques (e.g. when the timeline filter is applied, subsequent keyword searches are restricted to the specified time frame). This allows users to input as much information as they can to narrow the search space.

- *Improving our understanding of searching behavior*: The findings in Chapter 6 validate the some of the observations that were made in Chapter 4. One hypothesis was that as the need to locate events is central to the three searching tasks, we expected to see the same search methods being used to complete each of the three common searching tasks. We found that similar search methods were used for each task. In fact, there were no significant differences in the usage any of the six search methods across the task types.

We also found that multiple search methods were used. On average 5.6 (s.d. 2.3) search strategies were used by each participant. In fact, users were quite creative in combining search methods to achieve their goals. Ten out of the twelve participants used one or more unique search sequences.

In Chapter 4, we found that one the factors that affected search strategies was memory (how much information was known about at event). In Chapter 6, we found that user goals were quite varied across knowledge categories (*Precisely-known*, *Something-known* and *Unknown*). In *the Precisely-known* category, we expected most of the search strategies to revolve around the *Keyword search* as we thought this would provide the quickest access to an event. However, this was not always the case as some participants felt that it required less effort to scroll through the list of folders or use the *Timeline browser* than bringing up the keyboard and typing a search string. In this category, participants would use the information at hand to directly access a target event using the least mentally, physically and temporally demanding search method. In the *Something-known* category, the participants would first reduce search space. The first strategy was to use the *Timeline browser* to narrow the search down to particular time period. Participants would typically begin by picking the year, then the time of year and would then narrow the search down to a particular month. The second strategy was to cluster similar events by searching for a common keyword. The third search strategy was to search for a pre-existing category. This essentially restricts the search space to the category. The fourth strategy was to think of a small set of potential targets and then to work through them until the target folder was found. In the *Unknown* category, the participants would quickly refresh their memory by quickly skimming through relevant portions of the photo collection before beginning a more focused search. The specific strategies that were used varied from skimming through the folders to jumping to various temporal locations. Although, the goals in the each of the three knowledge categories were quite clear, we found that the search methods that were used to execute these goals were quite varied between users due to cognitive styles and abilities, user expertise and context availability.

- *A set of guidelines for designing photo search interfaces*: Our framework outlines design goals based on how much is known about an event (*Precisely-known*, *Something-known* and *Unknown*). These guidelines are based on the strategies that were recorded in each of the knowledge categories. When events are well-known, the search strategy is to minimize the time and effort required to retrieve them. The goal is to minimize both. One way of doing so is by designing a set of highly specialized and fined tuned search techniques that allow people to use the information at hand to directly access an event. As events are less well-known, the search strategy is to reduce the search space and then look for target events within the reduced set. Integrating multiple search methods allows users to input as much information as they can to narrow the search space. When events are unknown, the search strategies are more explorative to support learning and discovery. For this category, the goal of the search strategy is to maximize the knowledge gained, while minimizing the amount of interaction. On way of doing so is by allowing users to categorize events based on semantic properties. For example, with our search tool users were able to search for a common keyword to group semantically similar events.

6.1.2 Outline

Section 6.2 describes the design of search interface. Section 6.3 outlines the experimental goals and procedure. Section 6.4 presents the results. Section 6.5 discusses the findings and presents a framework that can be used to guide the design of future photo search interfaces. Lastly, Section 6.6 summarizes the major findings in this chapter.

6.2 Search tool implementation

The search tool system integrates four types of search techniques.

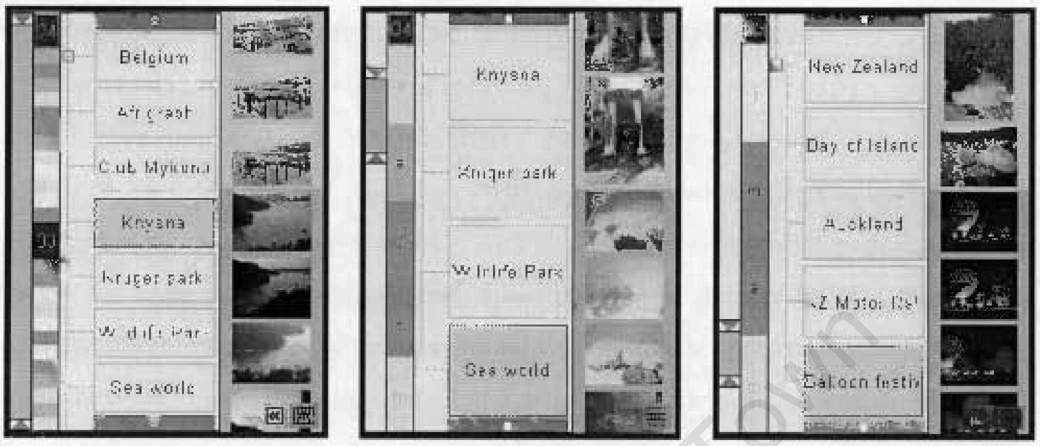
1. *Temporal browser* consisting of a *Timeline browser* and a *Timeline filter*.
2. *Hierarchical folder-based browser*.
3. *Keyword search*.
4. *Visual photo search techniques* (*AutoZoom* and *ManualZoom*).

These search techniques were not chosen arbitrarily and were based on recommendations from Chapter 4. By placing events on a timeline we preserve the natural order in which they occur. Each event can be identified by its unique temporal location (assuming of course that main events and sub-events are viewed as single events). The *Timeline filter* provides a way of narrowing the search space when the temporal location of an event can not be specified precisely. The *Hierarchical folder-based browser* is equally important. Folders are used to delineate events. Folder labels are used to store information about the event. The hierarchical structure is used to maintain relationships between groups of events. As users are familiar with hierarchical folder navigation, we are able to take advantage of pre-existing mental models and user expertise in navigating through these structures. The *Keyword search* allows users to search for any metadata that is encoded in folder labels. When keywords are known, it provides almost instant access to an event. The visual photo search techniques provide a way of visually searching through the photographs once an event has been found. The techniques can also be used to skim through a series of events. The next section presents each search method in isolation, before discussing how they are integrated.

6.2.1 Temporal browser

Figure 6.1 illustrates the *Temporal browser*. It is made up of two components, the *Timeline browser* and the *Timeline filter*. The *Timeline browser* is automatically generated using date information from each image. This date information is extracted from image EXIF headers when images are imported into the library. The import process takes place at start-up.

The *Timeline filter* is used to dynamically expand and contract the timeline. By doing so, it can be used to filter out parts of the timeline. Any years or months that are not visible on the screen are automatically filtered out. For example, Figure 6.1a shows the application in its initial state. The entire timeline is visible on the screen and no filtering is applied by default. In Figure 6.1b and Figure 6.1c the *Timeline filter* has been used to shorten the timeline, showing only a few months (May to September 2004 for Figure 6.1b and March to June 2003 for Figure 6.1c). Only these time periods are accessible, the rest of the timeline is filtered out. Of course this action is reversible by re-adjusting the *Timeline filter*. The *Timeline filter* is controlled by dragging the markers at the top and bottom of the widget.



(a) (b) (c)

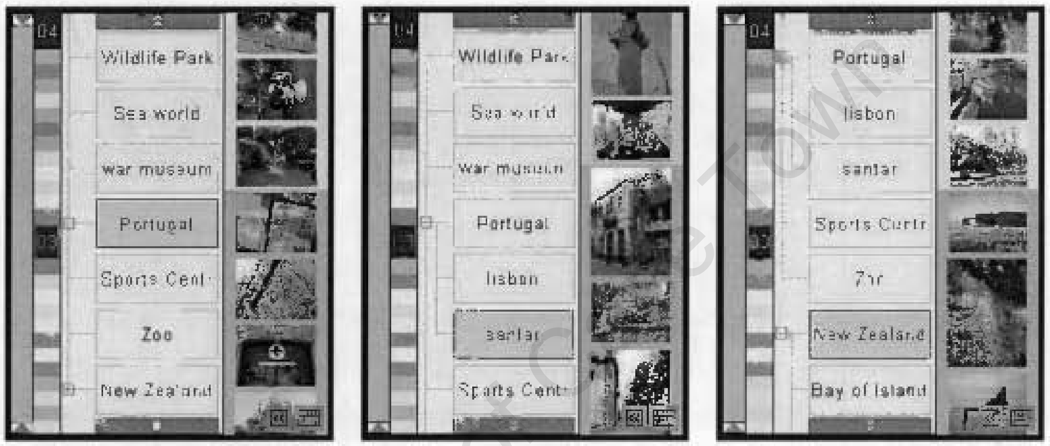
Figure 6.1: ***Temporal browser*** **consisting of the** ***Timeline filter*** **(left most control widget) and the** ***Timeline browser*****. In Figure 6.1a the** ***Timeline filter*** **is not applied, so entire time period is shown, in Figure 6.1b the** ***Timeline filter*** **has been used to restrict the photo collection to show pictures from May to September 2004 and in Figure 6.1c from March to June 2003.**

The *Timeline browser* provides a way of directly accessing a year or month. For each year, it displays a year tab and the months beneath it, in reverse order from December to January. The first letter of each month is used as a label to distinguish months (see Figure 6.1b and Figure 6.1c). Months are delineated using alternating grey bands. The blue bands are used to show months that have pictures associated with them. Two different shades of blue are used to differentiate consecutive months that have pictures associated with them. The current month is highlighted in orange. The current year is highlighted with an orange border. Years and months can be selected by tapping on them.

6.2.2 *Hierarchical folder-based browser*

Figure 6.2 illustrates the *Hierarchical folder-based browser*. The folder labels and structure is extracted from each photo collection. The functionality of this technique is similar to Window Explorer. The major difference is that the folders are always displayed in a single list and the hierarchical structure is encoded by indenting the connecting lines. This was done to show as much of the folder name as possible. Only seven folders were shown to reduce clutter and to account for the fact that people can only retain 5-9 items in their short term memory [80].

The 'expand' and 'contract' icons are used to expand and collapse folders. The buttons on either side of the list of folders are used to navigate up and down the list of folders. A folder can be selected by tapping on it.



(a) (b) (c)

Figure 6.2: ***Hierarchical folder-based browser.*** **Figure 6.2a shows that the 'Portugal' folder contains subfolders. In Figure 6.2b the 'Portugal' folder is expanded to show its sub-folders, 'Lisbon' and 'Santar'. Figure 6.2c shows the result of the using the navigation buttons to scroll down the list of folders.**

Figure 6.2a shows that a folder called 'Portugal' has been selected. This is visible from the highlighting and the blue border. The expand/contract icons serve two purposes. Firstly, they show which folders contain subfolders. Secondly, they can be used to expand and contract folders. From Figure 6.2a we can see that the 'Portugal' and 'New Zealand' folders contain sub-folders. In Figure 6.2b, the 'Portugal' folder has been expanded to show its sub-directories, 'Lisbon' and 'Santar', with 'Santar' being the currently selected folder. In Figure 6.2c the navigation buttons have been used to scroll down the list of folders. The arrows on the buttons indicate that you can scroll down further to access more folders. The square icon (see Figure 6.2a) indicates that you cannot scroll any further.

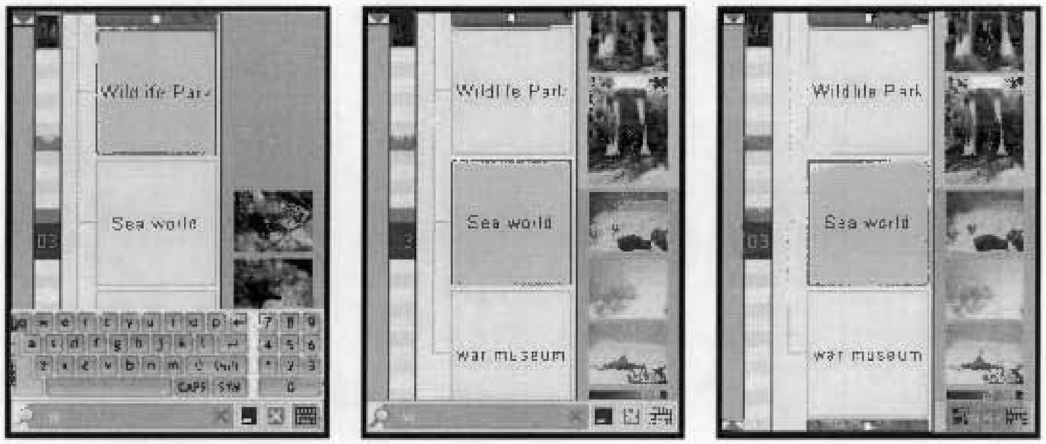
The pictures that correspond to the currently selected folder are highlighted using a dark grey band and are shown at the center of the screen. A few images from the next event are used to show happened next. Selecting an image from the next event automatically selects the corresponding event folder.

6.2.3 Keyword search

A dynamic letter and keyword search is implemented using a Pratricia-trie data structure [84]. This structure provides an efficient way of storing and retrieving alpha-numeric strings. The Pratricia-trie data structure is populated using keywords from folder names. The keywords are extracted using a rule-based approach to speed up the extraction process. For example, keywords are often delineated by using special characters (e.g. 'Wildlife Park' or 'Wildlife.Park' or 'WildLife_Park') or by capitalization (e.g. 'SeaWorld' or 'WarMuseum'). By applying various annotation templates, we can quickly extract keywords from the folder labels. For each folder label, the entire label is also entered as a keyword. This is necessary to ensure a match is always found when the entire label is inputted as a search string.

Figure 6.3a shows the components that make up the dynamic search. These are the keyboard, search box, clear search icon, minimize/maximize search icon, close search icon and the toggle keyboard icon. The keyboard icon on the far right is used to show or hide the keyboard as shown in Figure 6.3b. The search interface can also be minimized as shown in Figure 6.3c. This enables other navigation techniques to be used while in search mode.

In Figure 6.3, when the letter 'w' is entered, three matching folders are found. These folders have keywords beginning in 'w'. If the letter 'o' was entered next, two of the folders would be filtered out and only the 'Sea World' folder would be shown. The search is dynamic, so the search results are displayed as the search term is entered. The search is also fully reversible. It is possible to return to a previous state by deleting letters or to the original state by clearing the search.



(a) (b) (c)

Figure 6.3: The *Keyword search*. Figure 6.3a shows the main search components: keyboard, search box, clear search, minimize search, close search, toggle keyboard. Figure 6.3b shows the application in search mode with the keyboard hidden. Figure 6.3c shows the application in search mode with the keyboard minimized.

6.2.4 Visual photo search techniques

The search tool supports both visual photo search techniques, *AutoZoom* and *ManualZoom*. Both techniques have been described in detail in Chapter 3 and 4. We have added a small feature to both techniques to ensure that selected images are always animated in to the center of the screen. This also applies when photographs are displayed at their maximum size as shown in Figure 6.4a. When any non-central image is selected, it is smoothly animated in to the center of the screen.

The *AutoZoom* and *ManualZoom* techniques can only be selected in the detail view shown in Figure 6.4. The detail view can be selected by tapping on the thumbnail view shown in Figure 6.1, Figure 6.2 and Figure 6.3. The thumbnail view is then smoothly expanded to fill the screen. Alternatively, you can select this view by using the expand/contract view icon next to the keyboard icon. The two icons to the left of that are used to select the *AutoZoom* or *ManualZoom* techniques. The interface defaults to the most recently selected technique.

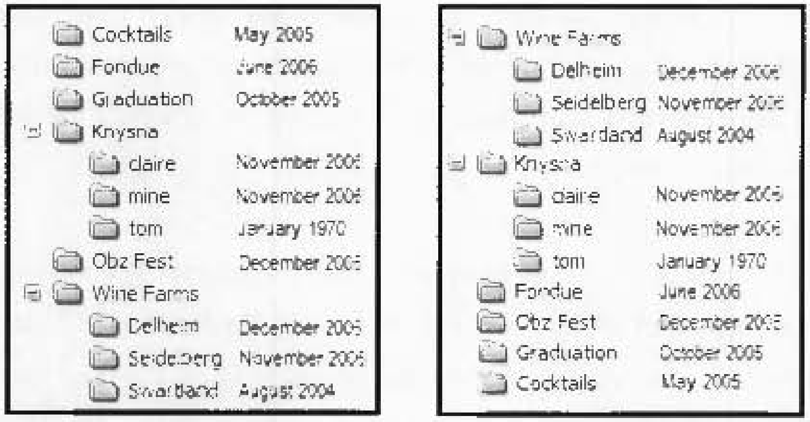
(a) (b) (c)

Figure 6.4: Visual photo search techniques. Figure 6.4a shows the detail view. You can select this view by tapping on a photograph in the any of the views shown in Figure 6.1, Figure 6.2 and Figure 6.3. Figure 6.4b shows the *AutoZoom* technique. Figure 6.4c shows the *ManualZoom* technique.

The current image is always shown in the center of the screen. The information bar at the top of the screen is used to show its position in a folder and also its path. For example, Figure 6.4a shows that the current image is the first image out of a total of 44 images in the 'Knysna' folder, making it the most recent image in the folder. It is also shows that its original path is 'Knysna\image1455.jpg.' The light and dark grey bands visible in Figure 6.4b and Figure 6.4c are used to show folder demarcations.

6.2.5 Integrating multiple search techniques

The first step was to find a way to integrate the timeline with the folder structure. The solution was based on two findings from Chapters 4 and 5. Firstly, users expect their folder structures to be preserved. Secondly, users expect events folders to be placed in the order in which they naturally occur. Within each folder, the photographs are sorted according the date on which they were captured. At each level of the hierarchy, the folders are sorted according to the most recent photograph they contain. Figure 6.5b shows the result of applying the date ordering scheme to an alphabetically ordered list of folders (see Figure 6.5a).



(a) (b)

Figure 6.5: (a) Alphabetical ordering and (b) our date ordering scheme.

(a) (b) (c)

Figure 6.6: Using perceptual cues to make relationships between multiple techniques more apparent when a date is selected (Figure 6.6a), a month is selected (Figure 6.6b) and photograph is selected (Figure 6.6c).

To be consistent, when a query is performed, the search results are ranked and displayed according to this scheme. The same ordering is also applied when visually searching through the images. When the *Timeline filter* is applied, the resulting action is reflected throughout the system. The folder structure only shows folders within this visible time period. The *AutoZoom* and *ManualZoom* techniques can only be used to locate images within this range. If a folder has photographs from a number of months, only the images from the valid time period are shown. The search is also restricted to this time period. Similarly, when a search is conducted, restrictions imposed by the search are also reflected throughout the system. The timeline only highlights the dates that are associated with the search results. The *AutoZoom* and *ManualZoom* techniques can only navigate through photographs in the results set.

In order to bring out correlations or disparities between the multiple views we had to ensure that when an action occurs along one dimension, the changes to other dimensions are immediately effected and made apparent. We used perceptual cues to make these relationships more apparent, focusing the user's attention on the relevant parts of the screen. For example, in Figure 6.6a when the user selects the month of June, it is immediately highlighted. The folder view also highlights all the folders contain photographs taken in June 2004. The most recent folder is selected by default and this highlighted by the blue border. The pictures from the selected folder are also shown on the screen, with the most recent photograph being shown in the center of the screen. In Figure 6.6b, the user selects the 'New Zealand' folder. The effects of this selection are immediately visible on the other dimensions. The timeline is of particular interest as it shows that the 'New Zealand' folder actually contains pictures from three months, February, April and May. The month of May is the most recent month and is selected by default to be the current month. In Figure 6.6c, the image in the center of the screen has just been selected. We can see the month and year in which it was taken in and also the folder in which it is stored.

6.3 Experimental Evaluation

The primary goal of the experiment was to validate some of the observations in Chapter 4. The hypotheses for this experiment were based on the following observations:

- ***The ability to locate events is central to all photo searching tasks.*** If this is true then we expect the same types of search methods to be used for each task type. Therefore, we hypothesize that the search technique usage across tasks is similar. We expect no significant difference in the usage of the *ManualZoom, AutoZoom, Timeline browser, Timeline filter, Keyword search* and the *Hierarchical folder-based browser* across task types.

- ***The search strategy used to locate an event depends on how much information is known about an event. Multiple search methods are needed to assist users in locating events.*** Based on the observations in Chapter 4, we hypothesize that the *Keyword search* will be used most when events are well known. We expect the *Timeline browser* and *Timeline filter* to be used more as events become less well known. We expect the *Hierarchical folder-based browser* to be used most when very little is known about a target event as users will scan through the list of folders to refresh their memory. In Chapter 4, the user's memory of an event was one of the more observable factors across participants that impacted search strategies. For example, when an event was known precisely, users where able to locate it directly by looking for particular event information. In contrast, when an event was unknown, users performed a more exhaustive search where they would often examine each potential target event. We hypothesize that user goals differ based on what is known about a target event.

- ***The AutoZoom and ManualZoom techniques are complementary in terms of how they support users in visually searching through photographs.*** In Chapter 4, we found that the *AutoZoom* technique is better for sifting through pictures, while the *ManualZoom* technique is better for inspecting pictures. As the search tool supports both techniques, we hypothesize that both techniques will be used together to take advantage of their respective strengths. Therefore we do not expect one technique to be used significantly more than the other when completing tasks.

6.3.1 Participants

Twelve participants took part in the experiment, five were female and seven were male. The participants were recruited through advertisements that were posted around the university. Five of the participants were undergraduates and seven were postgraduates. They were drawn from a variety of technical and non-technical backgrounds. While this sample may not be representative of the population as a whole, the range of differences that were observed helped us gain a fairly deep understanding of photo search strategies.

To effectively evaluate the search tool (and hence clarify the observations made in Chapter 4) the experiment was conducted using larger photo collections than the previous experiments. The participants that were chosen for this experiment were all early adopters that had fairly large collections. The mean number of photographs was 5134 (s.d. 2612), with a minimum of 2379 and a maximum of 8673.

6.3.2 Method

An observational study was used to gather data to clarify the observations that were made in Chapter 4. The 'Think Aloud' protocol was used to get the participants to describe their actions and their reasoning while performing the searching tasks outlined in Section 6.3.4. To aid this process, the searching tasks were designed to encourage social interaction with the evaluator. The entire experiment was captured using video. The video data was used for data analysis. Post-testing interviews were used to gain further insights into user behavior and decision processes.

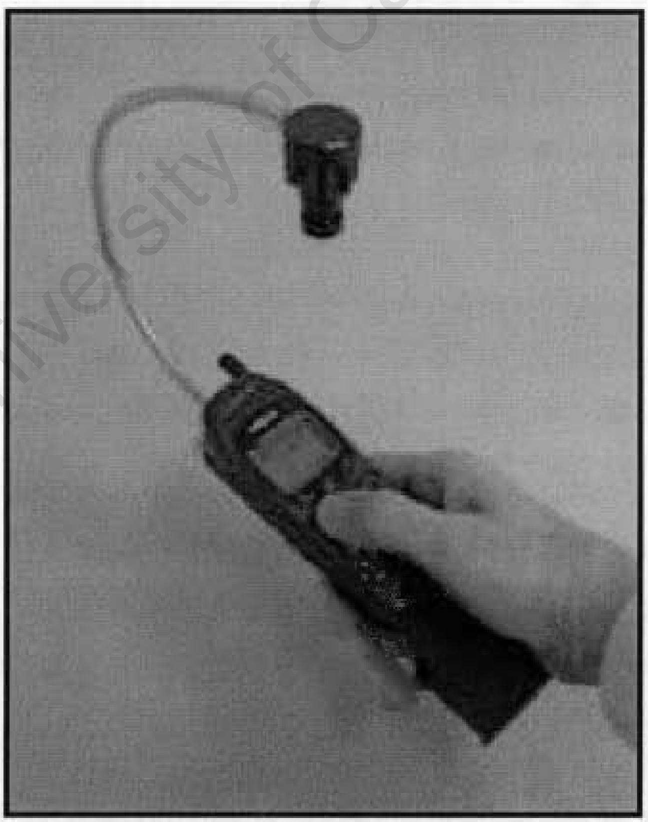


Figure 6.7: The mobile device observation camera that was used for the experiment.

6.3.3 Materials

The search tool was developed in C++ using a 2D graphics library called GapiDraw. It was deployed on a HP IPaq PocketPC 4100 series. A video camera was used to record all interaction on the device (see Figure 6.7). The audio was also captured using a microphone. Both feeds were combined to generate a video for each participant.

6.3.4 Procedure

Prior to the experiment, the participants were requested to submit their entire photo collection. Each collection was processed by resizing the pictures to the same resolution as the screen. This was necessary to ensure that thousands of photographs could fit on the 1GB SD card used in the device. Other than the resizing, the photo collections were not altered in any other way.

On arrival, the participants were told that they would be evaluating a photo search application for small display devices. A conceptual model extraction was conducted to ensure the participants were familiar with all the functionality before beginning the main experiment. The participants were then given a 10 minute break.

For the main experiment, the participants were required to complete 12 tasks which involved searching for *events*, *singles* and *properties*. For each of the three tasks types, the questions were precise, vague, verificative and exploratory. Of course these categories are not mutually exclusive. For example, a task can be both precise and verificative or vague and exploratory. Instead of using them as strict categories, they were used more as guidelines for probing different informational needs. The tasks were also tailored for each participant to ensure that they were as relevant as possible. An example of a sample question set is shown below.

Events

- *Precise*: Can you tell me about your trip to Seattle?
- *Vague*: What did you get up to on your trip overseas?
- *Verificative*: Were you at this year's Float party?

- *Exploratory*: I'm thinking about going away for the long weekend, can you recommend any trips that you have been on?

Singles

- *Precise*: You went to New York. Can you show me a picture of Trump Tower?
- *Vague*: You have just installed an instant messaging program on your phone and now you need to find a picture that your friends can identify you by. Can you show me one?
- *Verificative*: Did you go up sky tower? Have you got any nice aerial shots?
- *Exploratory*: Can you show me how you would pick out a wallpaper (or background image) for your phone.

Properties

- *Precise*: Your friends have just heard that you have got a new partner. They want to see some pictures. What would you show them?
- *Vague*: Are there any recurring events or hobbies that you take part in?
- *Verificative*: Do you have any special collections of scenic or artistic pictures in your collection?
- *Exploratory*: You are creating a collage for your best friend. Show me how you would go about finding suitable pictures.

The tasks were designed to initiate discussions and encourage the participants to show and tell. This was essential as we did not want the participants to know that we were interested in their search strategy, as this might influence their search strategy. The discussion was also used to gain some insight into thought processes. For example, participants might say; 'Oh let me show you this' or 'Where was that again' or 'I remember seeing that somewhere.' All these statements were invaluable in understanding search strategies.

A within groups study was conducted. The participants were split into six groups to balance the order in which the three tasks types were conducted to minimize learning effects.

An informal interview was conducted after the main experiment. The interview was used to clarify some observations and to distinguish elements of the searching behavior that were due to interface artifacts, as opposed to being part of wider searching behavior (see Appendix D for experimental materials).

6.3.5 Data captured

Each video was analyzed to extract quantitative and qualitative data for each task. In order to distinguish search strategies the following quantitative data was extracted:

- *Order:* The order in which search techniques were used to complete a task.
- *Time:* The amount of time spent using a search technique to complete a task.
- *Usage Count:* The number of successive times a search technique was used before proceeding to another technique

For each task, the qualitative data consisted of a short description explaining why the participant chose to search in the way that they did. We noted down how much information was known about the target event. There were three knowledge categories: *Precisely known*, *Something-known* and *Unknown* [76]. In the *Precisely-known* category, users have a specific event in mind. They are able to articulate what they are looking for and have a good idea of where to begin searching. In the *Something-known* category, users have some idea of the event they are looking for. They might not be able to articulate it or even know where to begin searching, but they are usually able to recognize it when they see it. In the *Unknown* category, users do not have any specific event in mind. They also have no idea where to begin looking. This data was obtained by collating user comments with our own observations.

6.4 Results

This section begins by describing search technique usage. It goes on to discuss usage patterns across participants and task types. Finally, it discusses search strategies in terms of the three knowledge categories (*Precisely-known*, *Something-known* and *Unknown*). When analyzing the data the list of search techniques was expanded by treating the *AutoZoom, ManualZoom, Timeline* and *Timeline filter* as separate techniques. This was done to obtain as much information as possible and develop a better understanding of search strategies.

6.4.1 Search technique usage

Figure 6.8 illustrates the number of times each search technique was used across the 12 participants: *ManualZoom* (MZ) = 125, *AutoZoom* (AZ) = 56, *Keyword search* (S) = 44, *Hierarchical folder-based browser* (F) = 121, *Timeline* (T) = 50 and *Timeline filter* (TF) = 2. The *ManualZoom* technique was used over two times more than the *AutoZoom* technique. The *Hierarchical folder-based browser* was also used over two times more than the *Keyword search* and the *Timeline browser*. The *Timeline filter* was only ever used twice. Although some techniques were used more than others, Figure 6.8 clearly indicates that different search techniques were used.

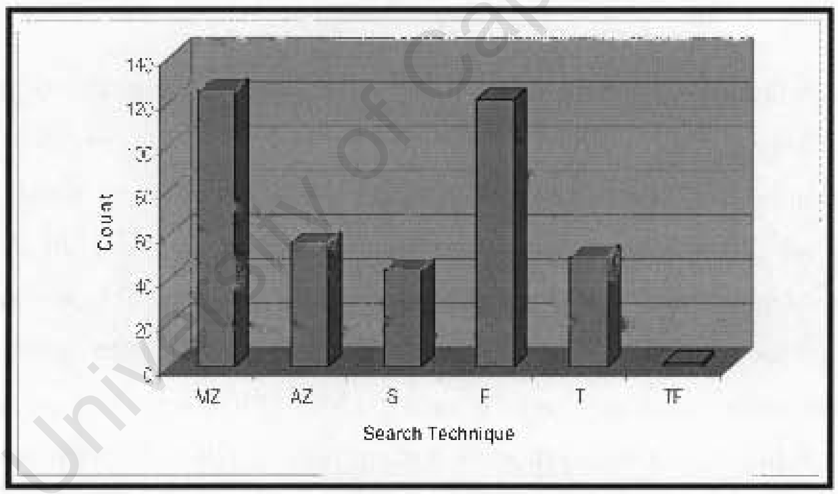


Figure 6.8: The number of times each technique was used in the experiment. *ManualZoom* (MZ), *AutoZoom* (AZ), *Keyword search* (S), *Hierarchical folder-based browser* (F), *Timeline* browser (T) and *Timeline filter* (TF).

Multiple search techniques were often used to complete tasks. Search techniques were sometimes used in different orders. Figure 6.9 shows how often search techniques were used in each position. It shows that the *Keyword search, Hierarchical folder-based browser* and *Timeline browser* were most likely to be used first. Following this, the *Hierarchical folder-based browser, ManualZoom* or *AutoZoom* techniques were most likely to be used next. From the second position onwards, there is a noticeable reduction in the number of times the *Keyword search* and *Timeline browser* were used. These finding are consistent with the temporal data shown in Figure 6.10. The data for Figure 6.9 and Figure 6.10 is presented in Table 6.3.

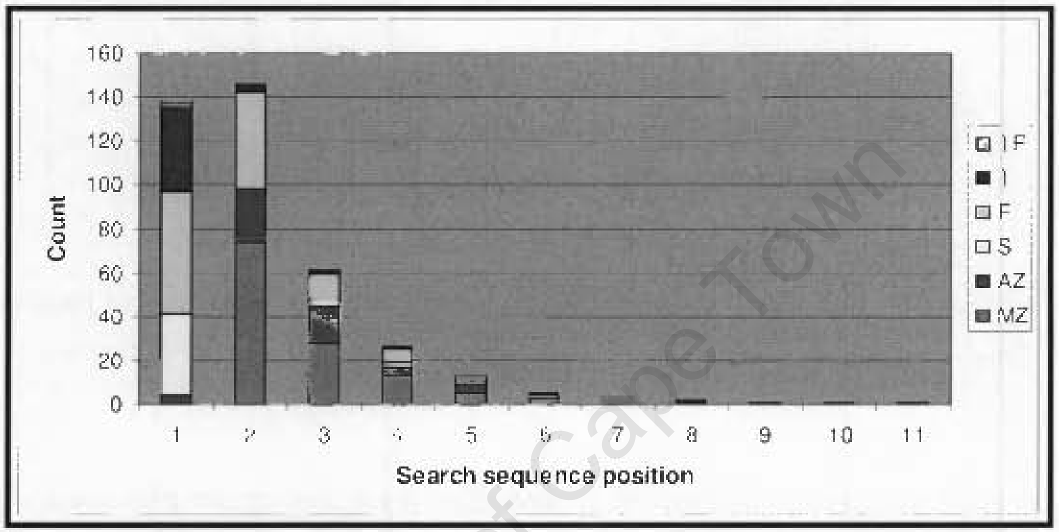


Figure 6.9: The search technique usage in different search sequence positions.

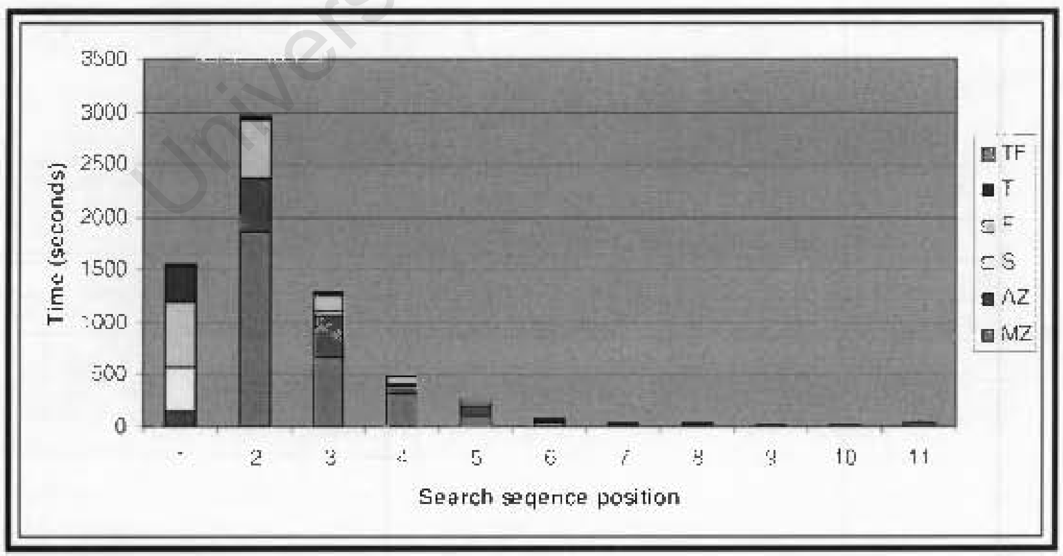


Figure 6.10: The amount of time search techniques are used in different search sequence positions.

	Usage count						Time (seconds)					
	Search technique						Search technique					
Position	MZ	AZ	S	F	T	TF	MZ	AZ	S	F	T	TF
1	0	4	37	36	39	2	0	150	416	629	343	16
2	74	24	0	44	4	0	1859	513	0	557	34	0
3	28	17	3	11	2	0	666	396	54	149	22	0
4	13	4	2	6	1	0	303	85	25	29	15	0
5	5	4	2	2	0	0	102	91	35	30	0	0
6	3	0	0	1	2	0	34	0	0	15	26	0
7	0	1	0	1	1	0	0	15	0	18	5	0
8	1	1	0	0	0	0	21	12	0	0	0	0
9	0	0	0	0	1	0	0	0	0	0	15	0
10	0	1	0	0	0	0	0	13	0	0	0	0
11	1	0	0	0	0	0	31	0	0	0	0	0

Table 6.1: The data for Figure 6.9 and Figure 6.10. It shows the number of times each search technique is used in each search sequence position. It also shows the amount of time that is spent using each technique in each search sequence position.

Table 6.2 shows all the search sequences that were used by the participants. For example 'F->MZ' specifies that the *Hierarchical folder-based browser* was used first and the *ManualZoom* technique was used second. For each search sequence its states the number of occurrences and the number of participants that used the same search sequence. The shortest sequence is one link long, while the longest is eleven links long.

Number	Search Sequence	Count	Number of participants
1	AZ	4	1
2	S->MZ	19	7
3	F->MZ	33	7
4	F->AZ	17	3
5	T->MZ	8	4
6	S->AZ	4	3
7	T->AZ	1	1
8	T->F->MZ	16	5
9	T->F->AZ	11	5
10	S->F->MZ	7	1
11	TF->T->AZ	2	2
12	T->MZ->AZ	1	1

13	S->F->AZ	[illegible]	1
14	S->AZ->MZ	1	1
15	F->MZ->AZ	[illegible]	1
16	T->MZ->F->MZ	2	2
17	F->AZ->F->AZ	2	2
18	F->MZ->F->MZ	3	2
19	F->MZ->S->AZ	1	1
20	T->F->AZ->MZ	1	1
21	S->F->AZ->MZ	1	1
22	F->T->F->MZ	1	1
23	T->F->MZ->F->MZ	2	2
24	S->AZ->MZ->S->MZ	1	1
25	S->F->AZ->S->AZ	1	1
26	F->MZ->F->MZ->S->MZ	1	1
27	S->MZ->S->T->F->MZ	1	1
28	S->T->F->MZ->AZ->T->AZ	1	1
29	T->F->T->F->MZ->F->T->MZ	1	1
30	T->MZ->T->F->AZ->T->F->AZ->T->AZ->MZ	1	1

Table 6.2: Search sequences used by the 12 participants

Figure 6.11 shows the distribution of search sequences based on the number of links they contain. The most popular search sequences are two links long.

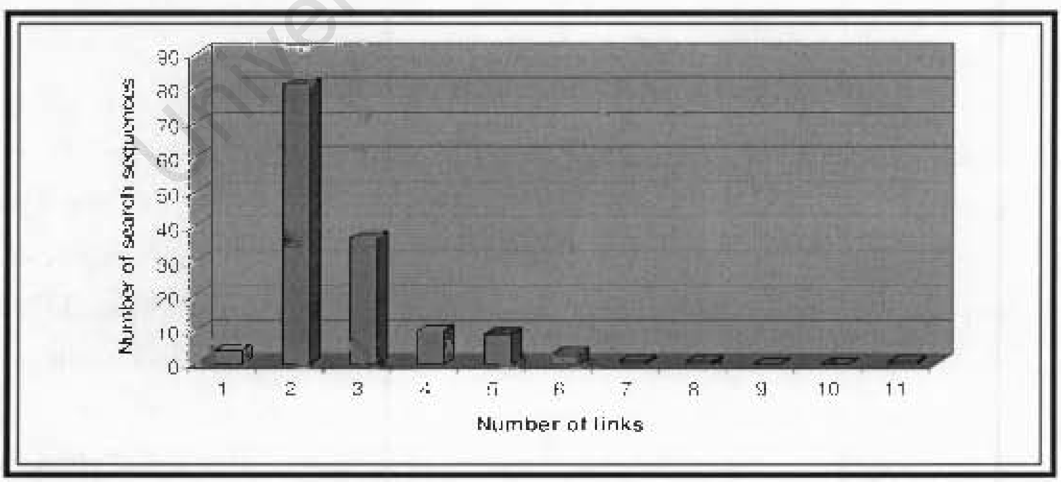


Figure 6.11: The distribution of search sequences based on the number of links they contain.

6.4.2 Search technique usage across participants

Figure 6.12 shows the search techniques that were used by each participant. For each participant, it also shows how often search techniques were used. On average each participant used 4.25 (s.d. 1.42) techniques.

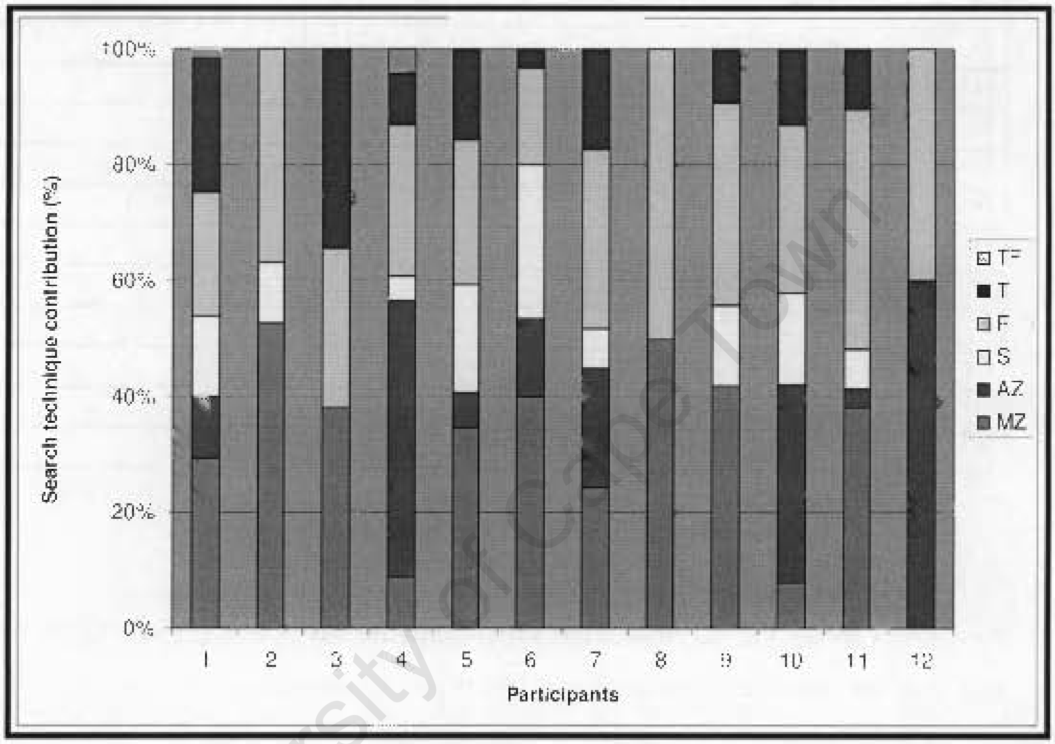


Figure 6.12: Search technique usage across participants.

Figure 6.13 illustrates the number of search sequences that were used by each participant. It also shows how often sequences were used. A different color is used to represent each search sequence. Figure 6.14 shows the search sequence ordered in terms of usage count. The same color scheme is used across participants to make easier to see whether or not a single search strategy is used predominantly more than the others. Some participants such as 2, 3, 8 and 12 predominantly use a single search sequence, while others use a variety of sequences. Figure 6.15 shows the number of participants that used each search sequence. On average each participant used 5.6 (s.d. 2.3) different search sequences. Ten out of the twelve participants used one or more unique sequences (see Figure 6.16).

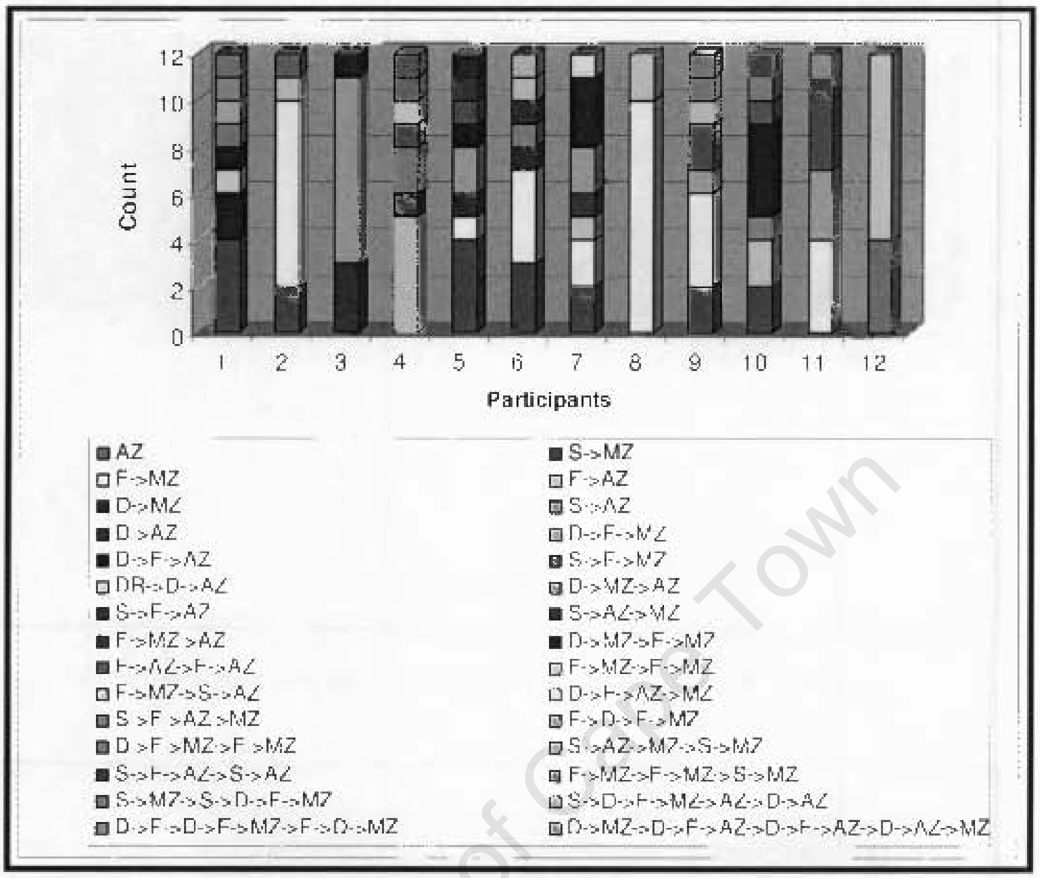


Figure 6.13: Search sequences used by each participant. Each color represents a unique search sequence.

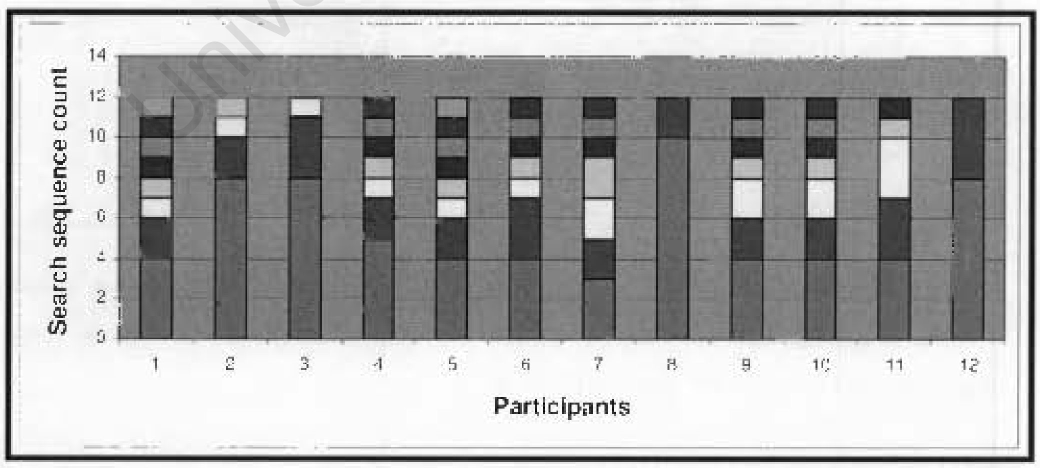


Figure 6.14: Search sequence usage across participants. Search sequences are ordered in terms usage. For each user, different colors are used to denote different search sequences. The same color scheme is used across participants.

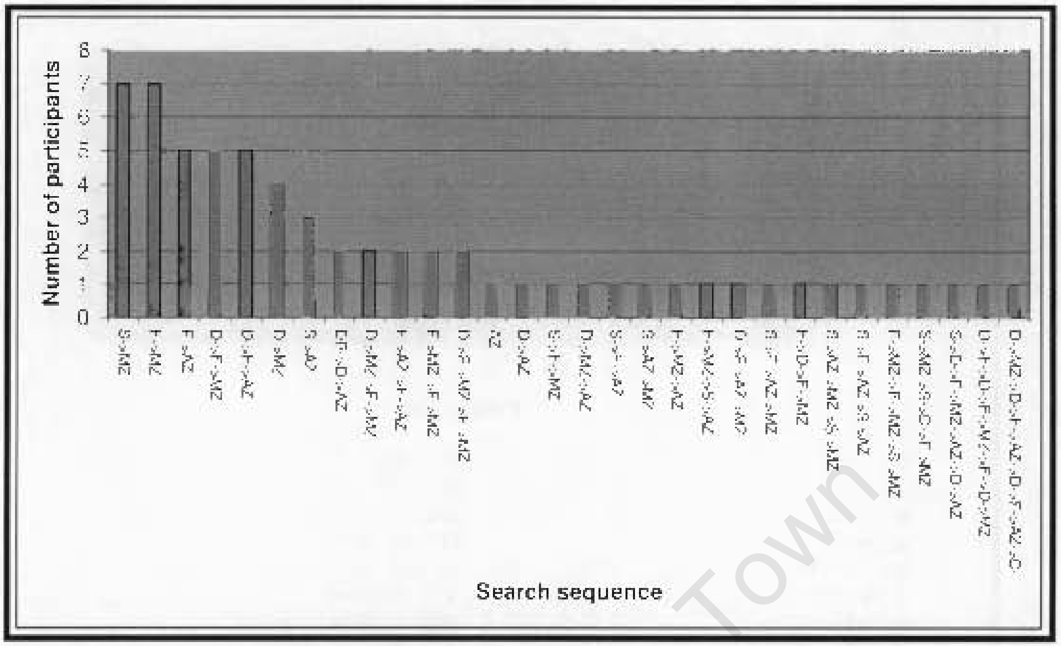


Figure 6.15: The number of participants that used each search sequence.

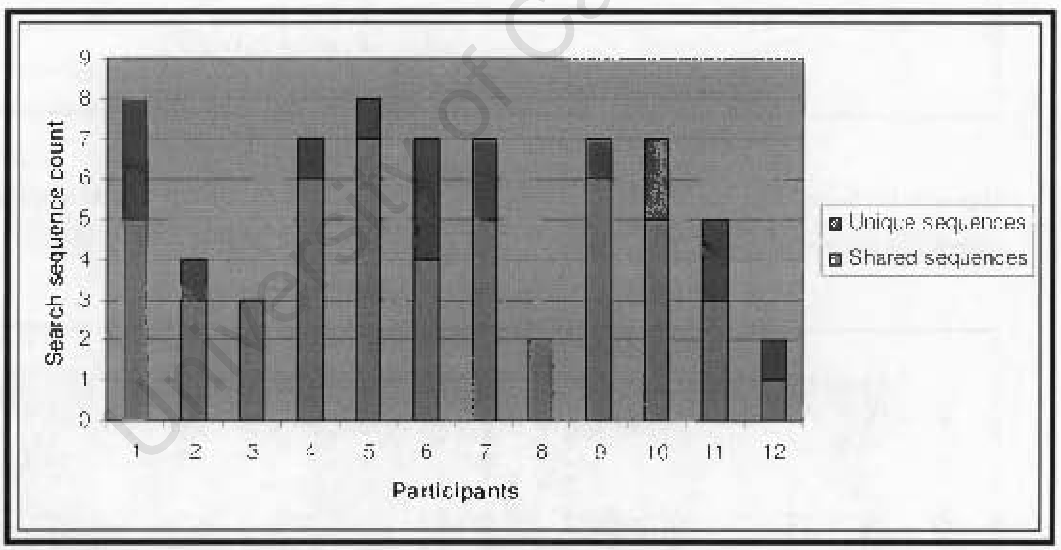


Figure 6.16: Search sequence usage across participants. For each participant it shows the number of search sequences that are unique to the user and also they number of search sequences that are used by other participants.

6.4.3 Search technique usage across task types

Figure 6.17 shows the distribution of search techniques across tasks types. For each task type it also shows how often it was used.

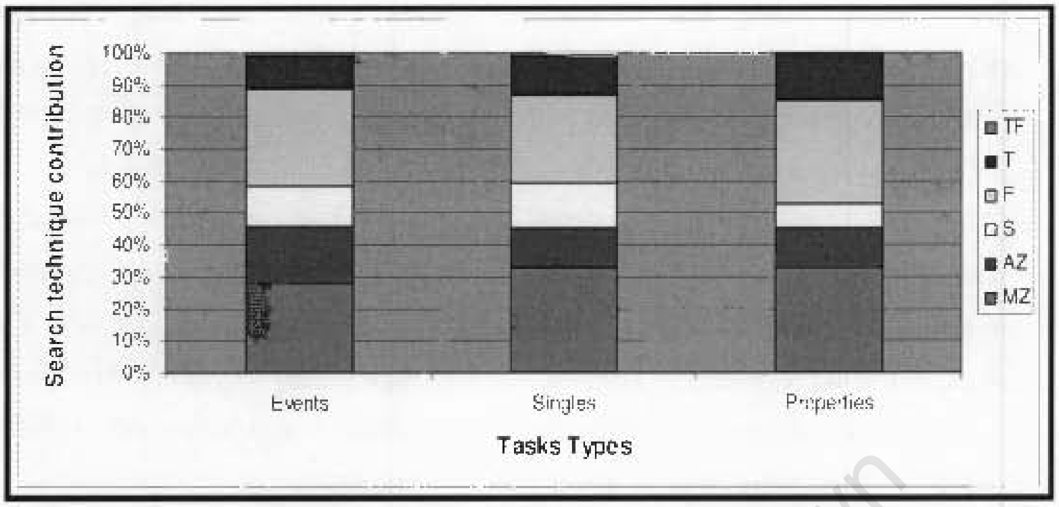


Figure 6.17: Search technique usage across task types.

For *Event* tasks, the *ManualZoom* (Kruskal-Wallis, KW=13.65812 p=0.0002), *AutoZoom* (KW=8.785831 p=0.0030), *Timeline browser* (KW=6.742495 p=0.0094), *Hierarchical folder-based browser* (KW=19.10966 p=0.0001) and the *Keyword search* (KW=8.801111 p=0.0030) were used significantly more than the *Timeline filter*. The *Hierarchical folder-based browser* was used significantly more the *Keyword search* (KW=9.401663 p=0.0022) and the *Timeline browser* (KW=10.72933 p=0.0011). The *ManualZoom* technique was also used significantly more than the *Keyword search* *(KW=5.264560 p=0.0218)* and the *Timeline browser (KW=5.868701 p=0.0154)*. There were no other significant differences.

For *Single* tasks, the *ManualZoom* (KW=13.70091 p=0.0002), *AutoZoom* (KW=5.311682 p =0.0212), *Timeline browser* (KW=6.738281 p=0.0094), *Hierarchical folder-based browser* (KW=19.06319 p=0.0001) and the *Keyword search* (KW=10.59517 p=0.0011) were used significantly more than the *Timeline filter*. The *Hierarchical folder based browser* was used significantly more the *Keyword search* (KW=5.828886 p=0.0158), the *Timeline browser* (KW=7.795366 p=0.0052) and the *AutoZoom* technique (KW=5.252703 p=0.0219). The *ManualZoom* technique was also used significantly more than the *Keyword search* (KW=3.815814 p=0.0508), the *Timeline browser* (KW=5.059720 p=0.0245) and the *AutoZoom* technique (KW=3.879647 p=0.0489). There were no other significant differences.

For *Property* tasks, the *ManualZoom* (KW=17.34857 p=0.0000), *AutoZoom* (KW=7.493213 p=0.0062), *Timeline browser* (KW=12.99767 p=0.0003), *Hierarchical folder-based browser* (KW=19.97186 p=0.0001) and the *Keyword search* (KW=7.493213 p=0.0062) were used significantly more than the *Timeline filter*. The *Hierarchical folder-based browser* was used significantly more the *Keyword search* (KW=14.22222 p=0.0002), the *Timeline browser* (KW=8.047263 p=0.0046) and the *AutoZoom* technique (KW=7.304149 p=0.0069). The *ManualZoom* technique was also used significantly more than the *Keyword search* (KW=8.831037 p=0.0030), the *Timeline browser* (KW=4.955536 p=0.0260) and the *AutoZoom* technique (KW=5.903649 p=0.0151). There were no other significant differences.

There were no significant differences in the usage of any of the search techniques across tasks (using the Kruskal-Wallis test at p=0.05). This shows that the task type is not a dominant factor in determining what search techniques are used.

6.4.4 Search strategies based on information needs.

Figure 6.18 illustrates how often the search techniques were used in each of the three knowledge categories. As you would expect, the *Keyword search* features most prominently when events are well-known and becomes less prominent when events are less well-known. In contrast, the usage of the *Timeline browser* is more prominent when events are less well-known. For each knowledge category, *Hierarchical folder-based browser* is used more than 20% of the time.

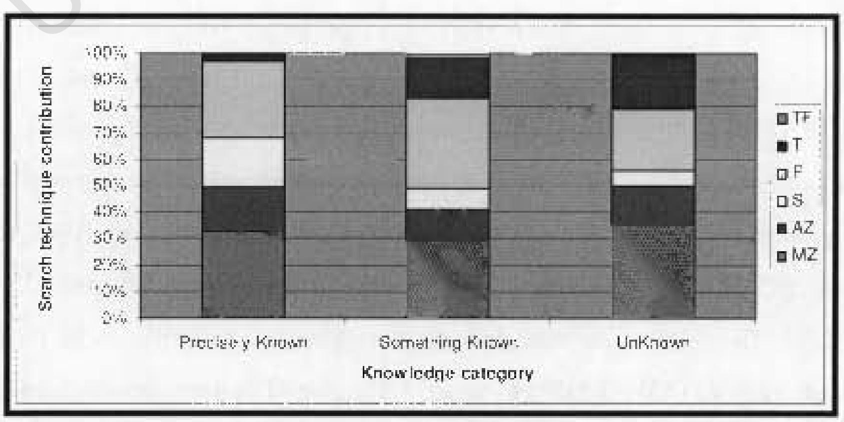


Figure 6.18: Search technique usage across knowledge categories.

Precisely-known knowledge category

To understand the significance of all the results presented above, it is necessary to have a closer look at the search strategies that are used in each knowledge category. Table 6.3 delineates the search sequences in the *Precisely-known* category into four main search strategies. It is important to remember that in this category events are well-known. Participants can articulate what they need and are able to formulate queries. Bearing this in mind, one would expect the *Keyword search* to be used most frequently. However, this was not the case (see Table 6.3).

Group	Search Sequence	Percentage	Group Percentage
1	F->MZ	24	46
	F->AZ	20	
	F->MZ->AZ	2	
2	S->MZ	29	38
	S->AZ	7	
	S->AZ->MZ	2	
3	T->F->MZ	5	12
	T->F->AZ	5	
	T->AZ	2	
4	F->T->F->AZ	2	4
	F->T->F->MZ	2	

Table 6.3: Four search strategies in the *Precisely-Known* knowledge category.

In Group 1, participants search for events by scanning through event folders for a particular event folder name. On finding the folder, they use the *AutoZoom* and/or *ManualZoom* techniques to locate target photographs. In Group 2, participants locate events by performing a *Keyword search*. Again, the *AutoZoom* and *ManualZoom* techniques are used to locate target photographs. In Group 3, participants locate event folders via dates. They know exactly when it occurred, so they select the date and then scroll to the event folder. In Group 4, participants start out by using the strategy in Group 1 and then switch to the strategy in Group 3. They realize that the first strategy will take too long and switch to what they perceive to be a quicker strategy.

Each strategy represents what participants believe to be the most cost effective in terms of time and effort (i.e. mental and physical demand). The latter being the reason why the search was not used more often. As these participants explain,

"I didn't use the search at all. I only have about 20 directories. It's far quicker for me to scroll through them than to bring up a keyboard and type a search string. I guess searching with a keyboard requires a lot more effort and precision than just scrolling, even if scrolling takes longer. If I had a phone with a keypad and predictive text, I think I would definitely use it more" [Participant 7]

"For me the date technique is the most accessible option. A single tap gets you to the desired location. Other methods require a bit more work. So I'll only use other methods when I can see a benefit of doing so or if the date technique is inadequate by itself." [Participant 1].

"It would be nice to have A-Z categories as well as the search. Although the search gives you the same features, it is not as efficient as just tapping one letter to see all the folders in that category. It's like using a shortcut. I'll always try to use the technique that requires the least effort. But if it takes too long, I'll consider using another method." [Participant 6]

As these comments suggest, two related factors that impact the search strategy are the device and the interface design. Both can impact how effective users are in terms of the time and effort required to complete a photo search task. There were other factors that affected the choice of a strategy such as, user expertise, cognitive styles (i.e. the habitual and preferred way of doing cognitive tasks [127]) and abilities (i.e. factors that contribute towards intelligence and influence the performance of specific tasks [60]). As this participant explains,

"I didn't use the search much. I was happy with scrolling up and down the list of folders. I guess I don't have many folders and it's the most familiar technique. It's the most similar to how I search on the desktop. One tends to be a creature of habit and you choose what you are

at ease working with and often don't look at other features. A couple of times I suddenly realize that it would have been so much easier to search. I guess with a bigger collection and lots more practice I would search more. It's not that I don't know how to use the search. It's just that my brain needs to register that its there. Part of that involves me figuring out when it's useful. Like I know it's useful when I know exactly what I want." [Participant 2]

The context availability limited the usefulness of some search techniques. As this participant explains,

"I don't really label my directories that well, so I can't make use of the search that much. I have to rely on the dates. I remember the months things happened in quite well. I remember when an event occurred and go straight to the month. I think I'd use the search and folder browser more if I had more meaningful labels. I think if I knew I had your photo browser I would put more effort into labeling my folders well. I guess I haven't put much effort into it because I haven't really seen the benefits." [Participant 3]

Tradeoffs were also made when choosing between the *AutoZoom* and *ManualZoom* techniques. On most occasions one technique was chosen over the other. Sometimes participants would flip between them (see Groups 1 and 2). As this participant explains,

"When I wasn't looking for detail the [AutoZoom] technique was better. It requires fewer movements which makes it better suited for browsing. So I would use the [AutoZoom] technique to get to a particular point, and [ManualZoom] technique to look closely at pictures." [Participant 4]

The *ManualZoom* technique was used at least two times more than the *AutoZoom* technique. The reason being,

"The [ManualZoom] technique has all the functionality of the [AutoZoom] technique with the added plus of extra control. The [AutoZoom] technique is nice when you want to scroll through a lot of pictures. However, the [ManualZoom] technique can also be used for that and it's better for checking out pictures. Its is better because you can search for global things

by zooming out all the way or you can look for detail by not zooming out as much and regardless of what you do you can make it move fast or not." [Participant 11]

Something-known knowledge category

Table 6.4 delineates the search sequences in the *Something-known* category into seven main search strategies. The most popular search strategy in this category (see Group 1) uses the *TimeLine browser* to narrow down the search and minimize the amount of scrolling. Once a suitable temporal position is found, participants scroll through the folders until they recognize the event folder. The *AutoZoom* and *ManualZoom* techniques are then used to locate target images. The more precisely a temporal location can be specified, the less scrolling is required. As this participant explains,

"The timeline was useful as a navigational tool because I don't know exactly when things happen, but I know the years and I know if it was early or late in the year. I used it more as a first pass thing, so I pick the year, then the time of year and then refine by scrolling through the folders. By searching this way you don't have many folders to scroll through. It's a good searching tool when you haven't organized your pictures as well as you would have liked to" [Participant 1]

We expected the *Timeline filter* to be more prominent in this category. However, there were a number of reasons why this was not the case.

"I didn't use the [Timeline filter]. I've only got three years worth of pictures. They are so widely separated already that I don't need to use it. If I had more pictures it would be useful because you would get to a point where it would be difficult to select a month or even a year. So the [Timeline filter] would be useful for filtering off irrelevant time periods to make the more relevant ones accessible. For me right now, all the years and months are accessible so it is quicker to just directly access them." [Participant 1]

"I didn't feel the need to use the [Timeline filter] because all my photographs are from the same era in my life. If I had 10 or 15 years worth of photographs I might use it to filter out different eras, like my high school era or my university era" [Participant 9]

"I don't know my photographs very well. If I filter some out, there is a chance that it might be something I'm actually looking for. So I prefer to be able to look through all my photographs" [Participant 7]

Group	Search Sequence	Percentage	Group Percentage
1	T->F->MZ	18.2	34.8
	T->F->AZ	10.6	
	T->MZ	3.0	
	TF-->T->AZ	3.0	
2	F->MZ	24.2	31.8
	F->AZ	7.6	
3	T->F->MZ->F->MZ	3.0	13.5
	T->MZ->F->MZ	1.5	
	S->AZ->MZ->S->MZ	1.5	
	S->F->AZ->S->AZ	1.5	
	F->MZ->F->MZ	1.5	
	F->AZ->F->AZ	1.5	
	F->MZ->F->MZ->S->MZ	1.5	
	F->MZ->S->AZ	1.5	
4	S->F->MZ	7.6	9.1
	S->F->AZ	1.5	
5	AZ	6.1	6.1
6	S->MZ	3.0	3.0
7	S->F->AZ->MZ	1.5	1.5

Table 6.4: Seven search strategies in the *Something-Known* knowledge category.

The next most popular strategy was to scan through the list of folders for target event folders (see Group 2). Participants were sometimes unable to describe what they were looking, but were able to recognize the target event folder when they saw it. Additional contextual cues, such as the image preview and date were used to confirm that they had found the correct folder. As this participant explains,

"Generally when I'm searching for photographs I do a scan. I find that the easiest thing to do is look through the folders until I recognize what I'm looking for. I use the dates to give me a

sense of which way I should be scrolling and also for confirmation as secondary step to me locating a folder. I don't tend to think of things in terms of time. It takes me more time to think of when something happened than to just scan through the folders" [Participant 8[

The search strategy in Group 3 is more exploratory in nature. This is because the participants were not able to narrow the search down to a single event. The search strategy involved investigating each possible target event folder. If all the candidate folders were known, participants would work through them from the most likely to the least likely. Otherwise, participants would work through them in the order in which they were remembered.

The search sequences in this group are characterized by inspections of more than one event folder. The differences in this group are due to how much is known about an event at the start of the task and also the knowledge that is accumulated during the task. For example, in search sequences 'F->MZ->F->MZ->S->MZ' and 'F->MZ->S->AZ' participants were able to learn and discover sufficient information to enable them to remember the name of the event they were looking for, allowing them to search for it.

In Group 4, the *Keyword search* was used to aggregate similar events, making it easier to compare them and search through them. As these participants explain,

"I sometimes used the search to cluster folders that have similar labels. For example when I search for 'Birthdays', I can aggregate all the birthday pictures together and then go through them." [Participant 9]

"The search is useful for clumping together events that I tend to do quite repetitively. Like I always go on December holidays, but sometimes I was unsure of which trip it was. With the search I was able to search for "December" and all December holidays pictures were grouped together in the search results. This made it easy to go through them and locate the one that I was after. It would be nice to extend this idea and have a search and keep facility. That way I can put different search results into a basket and then go through them to find what I want." [Participant 7]

In Group 5, target events and photographs were found by visually searching through images. This search strategy was only used by one participant. The participant felt that less effort was required to scroll through the images than navigating in and out of folders. As this participant explains,

"I like being able to visually search through images. The [AutoZoom technique] allows you to rapidly scroll through images without having to pay too much attention. Even when you are scrolling to fast to see individual photographs you can still make out events from the shapes and colors. I like the fact that it lets you scroll from one folder to the next without needing to select a folder." [Participant 12]

In Group 6, repeated searches were conducted to try and guess the event folder name. As this participant explains,

"When I know enough about the event, I'll try searching for all the possible names I can think of. It's worth a try. I've got a lot of folders and it takes a while to scroll through them." [Participant 6]

The search strategy is Group 7 takes advantage of the way people organize their photo collection. The *Keyword search* is used to locate a pre-organized category. This essentially restricts the search space to the category. Participants can then search through the category for target events and photographs. As this participant explains,

"I put a lot of effort into organizing my pictures in various categories. For example, I have a 'Weekend Away' category, 'Wedding' category, 'Birthdays' category. I will normally try to add new events to an existing category. If it doesn't fit I just create a new one. So usually I'll search for a category and then look through the folders. I don't remember individual folder names too well, but the category names are quite well ingrained in my head."
[Participant 10]

Unknown knowledge category

Table 6.5 delineates the search sequences in the *Unknown* category into four main search strategies. We found that the majority of participants would conduct an exhaustive search (see Group 1). Typically a date was selected first. This was a quick way of selecting the most recent event folder. Participants would navigate through the list of folders, checking each in turn. At the folder level, the image preview was used to make decisions on whether or not to investigate further. The *AutoZoom* and *ManualZoom* techniques were used to carry out more thorough inspections.

In Group 2, participants began by rapidly scanning through the entire photo collection to get a feel for what is in their collection. If the target event folder was not found in the initial phase, participants would then revisit event folders in order of the most likely to the least likely.

In Group 3, temporal navigation was used as a mechanism for jumping quickly to various locations in the photo collection. At each of these locations, participants would inspect any promising folders.

Group	Search Sequence	Percentage	Group Percentage
1	D->MZ	22.2	60.2
	D->F->MZ	5.6	
	D->MZ->AZ	5.6	
	D->F->AZ->MZ	5.6	
	D->MZ->F->MZ	5.6	
	D->F->MZ->F->MZ	5.6	
2	F->MZ	16.7	27.9
	F->MZ->F->MZ	5.6	
	F->MZ->F->MZ->S->MZ	5.6	
3	D->F->D->F->MZ->F->D->MZ	5.6	11.2
	D->MZ->D->F->AZ->D->AZ->D->AZ->MZ	5.6	
4	S->MZ->S->D->F->MZ	5.6	11.2
	S->D->F->MZ->AZ->D->AZ	5.6	

Table 6.5: Four search strategies in the *Unknown* knowledge category.

In Group 4, participants began by trying keyword-based searches. They felt that it was worth a try as it would save them from having to conduct an exhaustive search. However, the search terms were too broad and vague to return any relevant event folders. In both search sequences, participants had to resort to using one of the other search methods.

6.5 Discussion

Considering the results in the light of the hypotheses noted in Section 6.3

- **The task type is not a dominant factor in determining the search technique that is used**. For each task type, we found that the *ManualZoom*, *AutoZoom*, *Timeline browser*, *Hierarchical folder-based browser* and the *Keyword search* were used significantly more than the *Timeline filter*. The *Hierarchical folder-based browser* was used significantly more the *Keyword search* and the *Timeline browser*. The *ManualZoom* technique was also used significantly more than the *Keyword search* and the *Timeline browser*. The search technique usage was fairly similar for each task type. There were no significant differences in the usage of any of the search techniques across tasks. This shows that the task type was not a dominant factor in determining what search techniques were used. One possible explanation is that people think of their photographs in terms of events, so regardless of the task type, a core activity is always going to be locating events.

- **Multiple search methods are used to complete tasks. The user's memory of an event has an impact on the search strategy. The user's goals depend on who much information is known about an event.** We found evidence confirming that if you provide users with multiple searching methods, they will indeed use them. This is highlighted by the fact that on average 5.6 (s.d. 2.3) search strategies were used by each participant. In fact, participants were quite creative in combining search methods to achieve their goals. Ten out of the twelve participants used one or more unique search sequences.

We found that the usage of search techniques varied across the knowledge categories. As expected, the *Keyword search* featured most prominently when events were well-known and became less prominent when events were less well-known. The *Timeline browser* featured more prominently as less was known about an event. However, there were two unexpected results. Firstly, we expected the *Hierarchical folder-based browser* to be used when little information is known about an event. This was not the case as the *Hierarchical folder-based browser* was prominent in all three knowledge categories. One explanation is that user familiarity and expertise with this technique might have played a role here as some users might have elected to stay with their habitual and preferred way of searching, as this participant suggests, "*I was happy with scrolling up and down the list of folders. I guess I don't have many folders and it's the most familiar technique. It's the most similar to how I search on the desktop. One tends to be a creature of habit and you choose what you are at ease working with and often don't look at other features.*" Clearly, another factor at play here is the photo collection size. With a small collection it might be easier to scan through the list the folders than perform a keyword search as this participant suggests, "*I didn't use the search at all. I only have about 20 directories. It's far quicker for me to scroll through them than to bring up a keyboard and type a search string.*" Of course, this also depends on the interface and the device and also how tightly coupled they are, as same participant explains "*I guess searching with a keyboard requires a lot more effort and precision than just scrolling, even if scrolling takes longer. If I had a phone with a keypad and predictive text, I think I would definitely use it more.*" The second unexpected result was the lack of use of the *Timeline filter*. Participants felt that this technique would have been most useful for filtering out irrelevant time periods (or different eras). However, they did not use the technique as they felt the time span of their photo collections was not large enough to warrant the use of this technique.

A major focus of the experiment was to analyze the search strategies that were used within each of the knowledge categories (*Precisely-known*, *Something-known* and *Unknown*). The goals in each category were quite varied. In the *Precisely-known* category we expected most of the search strategies to revolve around the *Keyword search* as we thought this would provide the quickest way to directly access an event. However, this was not always the case as some participants felt that it required less effort to scroll through the list of folders or to use the *Timeline browser* than bringing up the keyboard and typing a search string. In this category the participants used the information at hand to directly access a target event using the least mentally, physically and temporally demanding search technique. In the *Something-known* category, the participants would first reduce search space. The first strategy was to use the *Timeline browser* to narrow the search down to particular time period. Participants would typically begin by picking the year, then the time of year and would then narrow the search down to a particular month. The second strategy was to cluster similar events by searching for a common keyword. The third search strategy was to search for a pre-existing category. This essentially restricts the search space to the category. The fourth strategy was to think of a small set of potential targets and then to work through them until the target folder was found. In the *Unknown* category, the participants would quickly refresh their memory by quickly skimming through relevant portions of the photo collection before beginning a more focused search. The specific strategies that were used varied from skimming through the folders to jumping to various temporal locations. Although, the goals in the each of the three knowledge categories were quite clear, the search methods that were chosen to execute these goals were quite varied between users due to cognitive styles and abilities, user expertise and context availability.

- **The *ManualZoom* technique was used more than the *AutoZoom* technique.** The *ManualZoom* technique was used significantly more than the *AutoZoom* technique. We expected the techniques to be used together, where the *AutoZoom* technique would be used to get to a particular point and the *ManualZoom* technique would be used to look closely at pictures. Most of the time, users chose one technique over the other. In fact, the techniques were only used together 9 times in the 144 search sequences that were analyzed.

> The main reason that was given for choosing the *ManualZoom* technique over the *AutoZoom* technique was that it *"had all the functionality of the AutoZoom technique with the added plus of extra control."*

The value of integrating multiple photo searching techniques becomes more apparent when looking at the number of search techniques that are used together in each category. In the *Precisely-known* group, 82% of the search sequences used two search techniques, while only 18% used three search techniques. In the *Something-known* category, 6.1% of the search sequences used one search technique, 40.8% used two search techniques, 50.1% used three search techniques and 3% used four search techniques. In the *Unknown* category, 44.4% of the search sequences used two search techniques, 33.3% used three search techniques, 16.7% used four search techniques and 5.6% used five search techniques. More search techniques were used as less was known about a target event. The fact that most participants in the *Precisely-known* category used two search techniques suggests that integrating multiple techniques as we did might not be the best solution for supporting this category. A better approach would be to allow users to customize the search tool by selecting the most appropriate search technique combinations for a particular task. This would allow more efficient use of the screen real-estate. Comparatively more, search techniques were used together in the *Something-known* and *Unknown* categories. Taking into consideration the searching strategies that were used in *Something-known* category, the value of integrating multiple search techniques is that the multiple inputs allow users to specify as much as they know, enabling them to narrow the search by placing constraints on multiple dimensions. For the *Unknown* category, multiple search techniques were used as users moved from one end of the knowledge spectrum to the other. The participants would quickly skim through the collection to rediscover the photo collection, find a starting point for the search and then begin a more focused search. Here, the value of integrating multiple search techniques is that it allows users to migrate from one end of the knowledge spectrum to the other, supporting them as their needs change.

Another interesting observation was the fact that the *Keyword search* was used to group together semantically similar events. The first method involved searching for a common keyword such as "trip". The second method involved searching for a pre-organized event category. This result is particularly interesting as it can open up new ways for locating events (or photographs) based on the type of information (i.e. who, when, where, what and why) and the nature of the information within these categories. The key is to this is to encourage users to be more consistent in their annotation practices when using this more light weight model for annotation (i.e. organizing photographs into events and labeling them). A large part of this involves making the benefits of doing so more apparent to the user, as this participant explains, *"If I knew I had your photo browser I would put more effort into labeling my folders well. I guess I haven't put much effort into it because I haven't really seen the benefits."*

The next section discusses the implication of these finding on design. It focuses on some of the main factors that impact photo searching strategies, that is, the domain knowledge, context availability, user expertise and cognitive styles.

6.5.1 Impact of domain knowledge

One way of thinking of the three knowledge categories (i.e. *Precisely-known*, *Something-known* and *Unknown*) is in terms of the knowledge gap. That is, the distance between the user's current knowledge of an event and the target knowledge needed to locate it. When an event is *Precisely-known* the gap between the current and target knowledge is minimal. When an event is *Unknown* the gap between the current and target knowledge is large. When *something is known* about an event, the gap may vary between these two extremes depending on the current knowledge. The challenge is how to assist users in bridging the knowledge gap.

Precisely known

When a user knows sufficient information to directly access a target event, a designer's challenge is to assist the user in making use of this information in the most effective way. In our study, we observed that some users were willing to compromise time for reduced effort. The goal should be to minimize both. The best way to support a user in this category is by designing a set of highly specialized and tuned search techniques. As a minimum, a searching toolkit should contain the following techniques:

- *Keyword search*: When users can articulate what they are looking for a keyword search provides almost instant access to an event.

- *Categories that support direct target acquisition and exploit spatial memory*: An example of this as an A-Z listing or Date listing that is directly accessible without any scrolling.

- *Active and passive links*: Provide mechanisms that allow users to 'actively' select things that they wish to remember. For example using tags, shortcuts, or favorites. Provide mechanisms that 'passively' remember with needing any user input. For example, dynamic lists of frequently accessed content or recently accessed content.

Unknown

When a user does not know what event they are looking for, it is usually because they have forgotten the event or can not think of any suitable event. In both cases users need to be exposed to the right information. In our study we found that a common strategy was to conduct an exhaustive search. However, this type of search can be time consuming and tedious. Some participants were more strategic by first getting an overview of the photo collection and then using this information to conduct a more focused search. The goal for designers is to assist users in meeting their initial needs, such as finding a starting point or learning about the information space. The design challenge is to maximize the amount of knowledge discovered while minimizing the amount of interaction. For example, one might use;

- Tag clouds: Flickr [35][46] allows user to navigate their photo collection using popular tags (see Figure 6.19). The more popular the tag the larger it is displayed. This idea could be extended for event folder labels. For example, keywords in folder labels might be 'boosted' based on a number of criteria, such as frequency, popularity, recency, size, or time span.

Figure 6.19: Flickr's photo navigation based on the popularity of a keyword.

- Photo summaries: Rother et al. [110] have developed a system that can be used to auto collage an entire photo collection into a single image (see Figure 6.20). Such an approach could assist users in learning and rediscovering their photo collection. It could also be useful for providing previews of photographs in each folder. Snavely et al. [117] have developed a system that summarizes a set of photographs by building a 3D model that supports 3D navigation and exploration (see Figure 6.21). It also estimates possible camera positions (see Figure 6.22). Instead of reviewing all the images, users would only need to select a few images from each vantage point to get a gist of what is contained in the photo collection.

Figure 6.20: AutoCollage software creates a photo summary of an event [110].

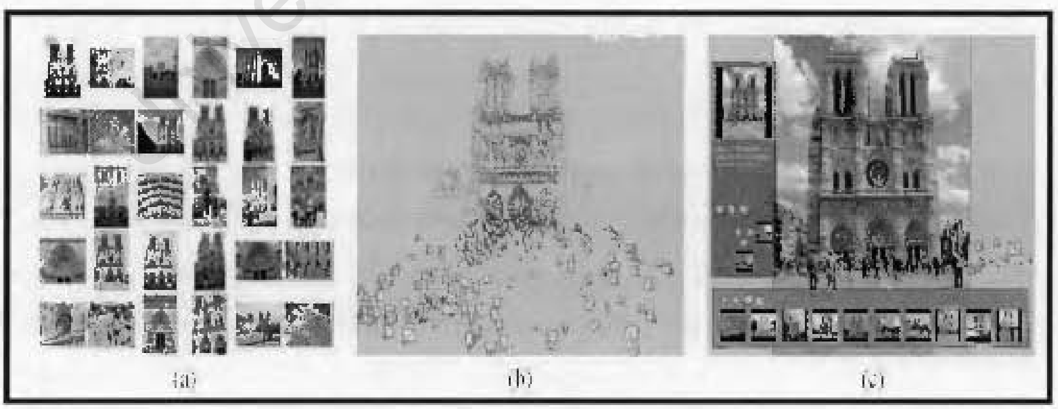


Figure 6.21: Building a 3D model out of a set of pictures. The system takes collections of photographs (a), reconstructs 3D points and view points (b) and enables 3D navigation and exploration of a set of images (c) [117].

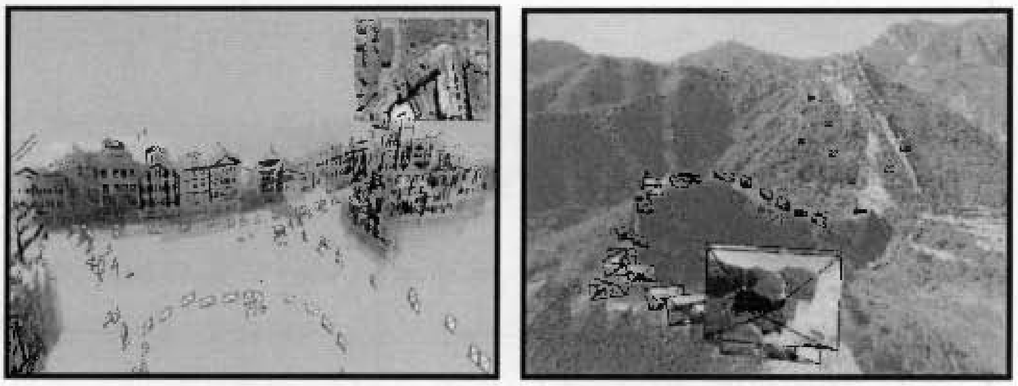


Figure 6.22: Estimating camera positions [117].

- 'Seeded' searches: Websites such as Flickr have highly categorized and structured image collections. A user might use such as service to find the kind of pictures they are looking for in their photo collection and then use this as starting point to request similar content in their photo collection. The system would then use the set of images and its associated metadata to search through their personal photo collection and return similar events.

Something known

This category is challenging to design for because the requirements for search tools differ according to how much is known. When the knowledge gap is large, users will require solutions that enable them to explore, discover and learn about their photo collection. As the knowledge gap narrows, users will require solutions that enable them to investigate. When the knowledge gap is small, users need solutions that provide immediate access to the information that they are seeking.

One must also consider the range of investigative search activities that users are engaged in. For example, we found that users were engaged in search activities such as: recall and recognition (e.g. scanning through a list of folders until a target is recognized); filtering (e.g. using temporal navigation to narrow the search space down to a specific year or month) and aggregation (e.g. grouping events that share similar keywords using the *Keyword search* technique); and verification and negation (e.g. inspecting multiple candidate folders). Furthermore, information about an event can be split along a number of dimensions (who, when, where, what and why). At the start of a task users may only be able to specify a limited amount of information along each dimension. As users learn and discover more information, searching tools must ensure that this information can be used in a way that complements what they already know. The searching tool must allow users to input any information that is available to them, allowing users to search simultaneously along multiple dimensions. The versatility of the integrated approach enables it to address these concerns. The seamless integration is key to supporting investigative activities. The multiple inputs allow users to specify as much as they know, allowing them to search in a more natural way.

A huge challenge with the integrated approach is the user interface design, as each technique needs to be made accessible, but at the same time great care must be taken not to overwhelm the user with too many options. One solution is to allow customization. This will allow users to select the techniques that can be integrated together, allowing them to optimize the solution to cater for device constraints as well as other cognitive and situational factors. It is unlikely that software can be built to seamlessly integrate every possible combination of search tools. Where search techniques can not be easily combined together, users should be able to choose different workable combinations, similar to how users can select preset equalizer combinations on MP3 players such as Winamp[2] to suit a particular genre of music.

6.5.2 Impact of context availability

Figure 6.23 illustrates the Context-Resource (CR) model. This model is used to help designers appreciate the impact of context availability on the navigation schemes they provide. Figure 6.24 represents the navigation in an application that is highly adapted for the environment. A great deal is known about the context and the application should help users get directly to the resources they need. For example a journalist might be writing an article on the Darfur region and needs some pictures. Instead of having to hunt through the digital archive, the photo application might

[2] http://www.winamp.com/

provide some pictures based on the content of the article and the annotations in the digital archive.

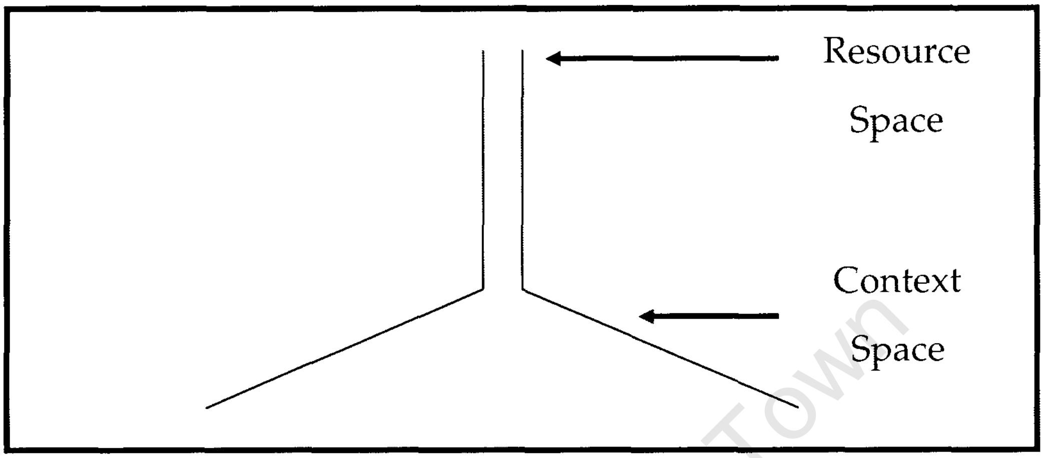


Figure 6.23: The Context-Resource model. When a great deal of user-context information is available the resource space should be small.

However personal photo applications have navigation styles more like Figure 6.24. Sparse or no context information is available. As a consequence, the user has a large information space they have to work their way through. Although new devices with access to user-context will reduce the number of applications using this navigation style, larger resource space traversal will continue to be important for two reasons. Firstly, even when there is a broad range of context available, the user might wish to narrow the context space to extend the amount of choice in the resource space (see Figure 6.25). For example, imagine you need lunch but you are in an unfamiliar part of town. Your handheld could recommend a few restaurants in your immediate vicinity which match your tastes and show you photographs of the restaurant or menu. If the restaurant finding application gives two unappealing options, for instance, the user might relax the location constraint, to look further a field. Secondly, users will always want to access information that is completely unrelated to their context, for example to rediscover events and pictures they might have forgotten about. Providing access to wide range of information will appeal to many people, especially in instances when people have a great deal of "dead-time" to fill.

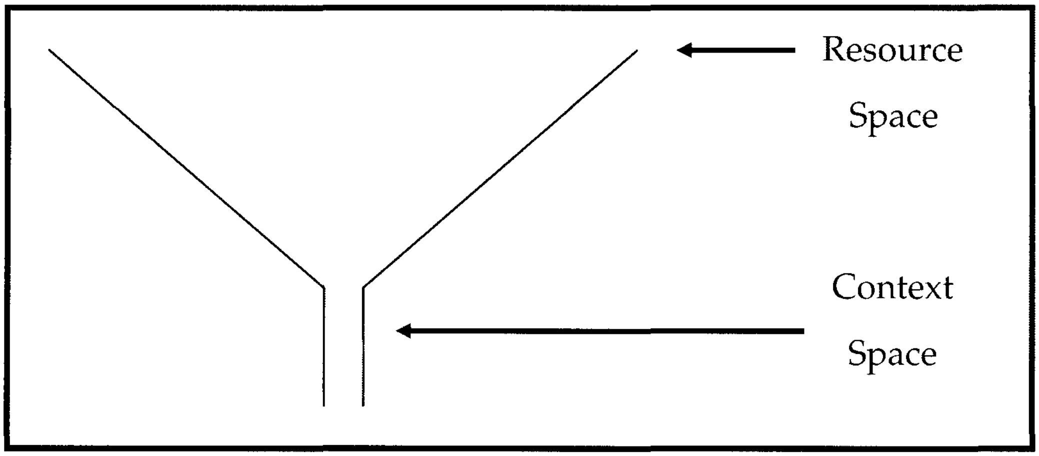


Figure 6.24: The Context-Resource model. When the context space is narrow, the resource space is large.

Where the resource space is large, the navigation technique must provide ways for the user to quickly get an overview and provide them with ways to reduce the search space by filtering so that they can get the details when required.

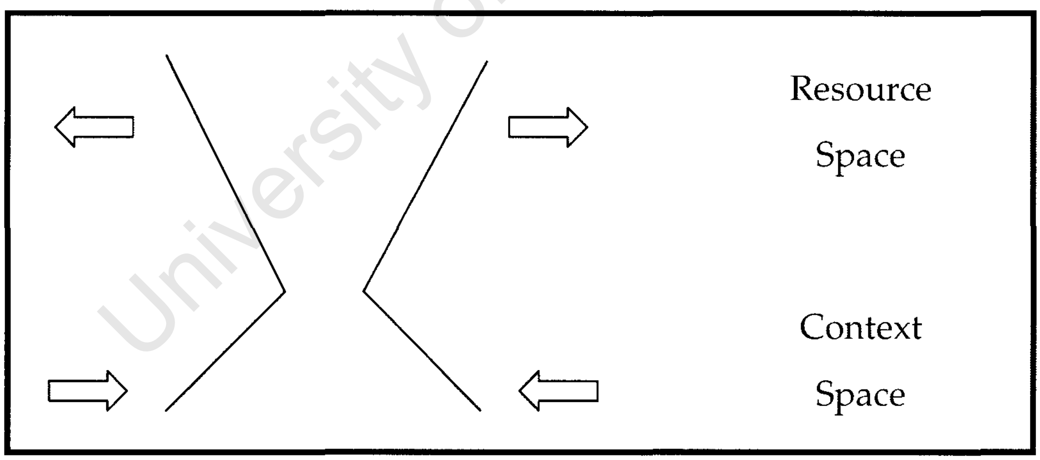


Figure 6.25: The Context-Resource model. As the context space is made smaller (by information availability or user choice), the resource space expands in size.

Symmetrically, where the resource search space is much more narrow, a good navigation style would seem to be, 'detail first, aggregate and overview.' For example, imagine you are interested in changing your wallpaper. A good place to look is in your 'scenic pictures' folder, where you always save your best pictures. You search for the folder and find a photograph that has some of properties you are looking for. You wonder if there are any more like it, so you then aggregate by event to see similar pictures from the same event. You still don't find anything suitable, so you aggregate by location to see if there are any other events from the same location. At a glance you can also see how events are distributed geographically.

To give the user as much flexibility as possible, the navigation technique should provide top down navigation (overview first, filter, details on demand) and bottom up navigation (details first, aggregate, and overview). Ideally the navigation technique should provide a seamless transition between the two. This will help support lookup, learning and investigation, especially in more exploratory contexts.

6.5.3 Impact of user expertise

The goal of the conceptual model extraction before the main experiment was to allow users to gain some experience in using search techniques and to show them how the techniques were meant to be used. Despite this training, we found that previous experience influenced user preferences for a particular search strategy. For example, most users were experts in using windows explorer to locate photographs, so some would still opt for the more familiar *Hierarchical folder-based browser*, even when the *Keyword search* was clearly a better choice. The challenge here is to make the cost of interaction (such as the physical, mental and temporal demand) more apparent. This way, users can choose the most cost effective technique.

As users became proficient with the various search techniques, they began looking for ways to optimize these techniques. In particular, participants wanted additional functionality to allow them to skim through the folders more quickly and also to expand and contract the hierarchical structures more easily. It is essential for users to be able to optimize the interface, especially when conducting repetitive actions. For example, by allowing users to create shortcuts or automating action sequences.

6.5.4 Impact of cognitive styles

In the study, we found that searching strategies were highly variable between individuals. A total of 30 unique search sequences were used by our participants. For each search strategy, there were often a number of different search sequences. One of the reasons is that people have varying cognitive styles (the habitual and preferred way of doing cognitive tasks) and abilities (intellectual aptitude at performing particular cognitive tasks). Studies on information seeking on the web [127][96] suggest that users with different cognitive styles develop different seeking strategies. According to Marchionini [75], although the basic cognitive processes (e.g. thinking, remembering and problem solving) and activities (e.g. recognizing, identifying, formulating, gathering, differentiating and verifying) are similar between individuals, the search tactics differ significantly because some people are highly tolerant of ambiguity and uncertainty, whereas others demand specificity and completeness. Photo applications must be designed to accommodate the diversity of cognitive styles. Supporting multiple methods within a single framework addresses this need as it allows users to navigate in a variety of different ways, choosing the methods that are most appropriate for them.

6.6 Summary

Section 6.1 motivates the importance of supporting multiple search strategies. Section 6.2 describes the design and development of a single search interface that integrates multiple search techniques. Section 6.3 outlines the experimental setup. Section 6.4 presents the results. Section 6.5 analyses the results and provides two frameworks that can used to inform the design of photo search interfaces. This section reflects of the major findings in this chapter.

The study was used to clarify the some of the observations that were made in Chapter 4. We found that the search technique usage was similar across task types. One explanation that was given was that people think of their photographs in terms of events, so regardless of the task type, a core activity is always going to be locating events. We also found that multiple search methods were used. On average 5.6 (s.d. 2.3) search strategies were used by each participant. In fact, users were quite creative in combining search methods to achieve their goals. Ten out of the twelve participants used one or more unique search sequences. As observed in Chapter 4, we found that search strategies did differ based on how well events are known. This was based on the fact that user goals were quite varied across the three knowledge categories. There were other contributing factors that influenced search strategies: cognitive styles and abilities, user expertise, and context availability.

We found that value of integrating multiple search techniques is that it provides the kind of versatility that is needed to accommodate these factors. The key however, is finding the right blend of search techniques that is needed for a photo search task, bearing in mind that user goals will change as they bridge the knowledge gap. A single configuration (or bundle) of search techniques is unlikely to suit all users. One solution is to support customization so that users can optimize a search tool for search tasks, device constraints, as well as other cognitive and situational factors. The key to supporting customization is to ensure that each technique is built around locating events. This makes it easier add or remove a search technique.

Chapter 7

Conclusions

Section 7.1 reviews the main aims of this thesis and the methodology that was employed. Section 7.2 discusses the main contributions of this work. Section 7.3 offers a critical reflection of this thesis. Finally, Section 7.4 presents a number of suggestions for further work.

7.1 Revisiting the research goals and methodology

This research is aimed at improving HCI knowledge in the design of the next generation of photo searching tools for small-display devices. Today, these devices have all the ingredients for a truly mobile photo collection (large storage, multiple networking capabilities and high resolutions screens). However, they lack the tools for searching through large collections of photographs [61]. This is particularly important as users have expressed a desire to store images on mobile devices in the long term [61].

No substantial research has looked at addressing users needs. Few researchers have considered the importance of supporting both searching and browsing to cater for these needs. None that we could find have assessed the potential impact of adding desktop-based annotation capabilities. Consequently, the aim of this thesis was to provide an empirical foundation for the design of photo search tools on small screen devices.

The objectives were defined as follows:

1. *To devise a set of empirically grounded guidelines for designing photo search interfaces on small display devices.*

2. *To develop a photo search tool that addresses the limitations of current systems on small display devices.* In contrast to previous work, the design of search tool takes into account the task types (*events, singles* and *properties*), the methods of information access (searching and browsing) and current annotation practices where photographs are organized by events and a description is provided for each event.

To achieve these objectives an iterative user-centered design methodology was employed. The end practical result was a single photo search interface that incorporates the best traits of a variety of tools to support search. The thesis reflected on each cycle in the iterative design process. In Chapter 3, an initial photo searching tool was developed based on the user needs analysis that was conducted in Chapter 2. The design was deliberately incremental so that the search tool could be benchmarked against a baseline thumbnail-grid browser. The emphasis was on improving current visual photo search tools by taking into account users needs when visually searching through photo collections. A comparative study was conducted to compare the techniques in terms of efficiency, effectiveness and satisfaction. The new systems were substantially better than the baseline system on all three measures and were found to be particularly good for inspecting photographs. The next iteration in the cycle (see Chapter 4) was aimed at identifying the problems with the search tool in terms of supporting the three common search tasks so that they could be addressed to provide a better search experience. We found that there was a need to provide more rapid access to events to support the photo search tasks more adequately. We found that multiple search strategies were used to locate events. The strategies were often influenced by the user's memory of a target event. In a follow up study in Chapter 5, we found that people predominantly organize photographs by events. We found that users would annotate each event using multiple types of information. For the final iteration in the design process (see Chapter 6), a search tool was developed based on the recommendations in Chapter 4 and the follow up study in Chapter 5. The search tool integrated multiple search methods (*Keyword search, Timeline browser, Timeline filter, Hierarchical folder-based browser, AutoZoom* technique and the *ManualZoom* technique) within a single user interface. The blend of search methods was chosen to allow users to search for events when they were well-known and also when they could not be recalled precisely. A study was conducted to see how well the search tool was able to support users in performing the three search tasks. The main outcome of the study was a framework for designing future photo search interfaces.

7.2 Contributions

The first contribution to field of Human-Computer Interaction improves existing knowledge on photo searching behavior by providing a number of empirically grounded findings about searching behavior. The second area of contribution is a single photo search interface for small display devices that is based on iterative studies of various user interface designs that has tried to incorporate the best traits of a variety of search tools to support the common searching tasks.

7.2.1 Improving knowledge of photo searching behavior

The research in Chapter 4 focused on identifying the problems with the initial photo search tool so that they could be addressed to provide a better search experience in terms of supporting users in performing the three common searching tasks. No previous research that we could find has focused on understanding how to support users in locating *events*, *singles* and *properties*. Based on an observational study, we found that when searching a photo collection the participants would primarily think of their photographs in terms of events. This seemed to occur irrespective of the task type. For example, when locating a *single* they would first try to associate it with an event, then locate the event and finally locate the target photograph. When locating *properties*, they would think of events that were likely to contain target photographs and then navigate from one event to the next. Often participants would automatically think of the date as secondary process to identifying a target event. When the date was precisely known, they would navigate directly to the date. Otherwise, they would narrow the search down to a year or month by thinking of when it happened relative to neighboring events or landmark events. The task of locating an event was made easier by the fact that people organize photographs by events. Participants noted that the metadata that is associated with an event encodes the kind of information that they are likely to remember when searching for an event. We observed that this metadata was used extensively when locating target events. Some participants noted that they preferred to use the folder labels to distinguish events rather than scanning through the photographs. We also observed that some participants would organize events according to special themed categories. When searching for an event in a category, the participants would first identify the event, then think of an appropriate category and finally navigate to the category. This particular search strategy was problematic with our initial search interface as the chronological ordering broke up the categories. When the spatial location was known and participants could not pinpoint an event temporally, we found that participants preferred to browse spatially rather than temporally as they felt it was more cognitively demanding to constantly resolve and reevaluate where you are temporally in relation to a target event. Based on these observations, it was evident that multiple search strategies were used to locate events.

There were a number of factors that influenced the search strategy. We found that for smaller collections users would often do an exhaustive scan to locate a target event as this required little effort and was not too time consuming. Participants with larger and older collections would often try to narrow the search down to a particular time period to avoid having to perform an exhaustive search. The user's memory of an event was one of the more observable factors across participants that impacted search strategies. For example, when an event was known precisely, users were able to locate it directly by looking for particular event information. In contrast, when an event was unknown, users performed a more exhaustive search where they would often examine each potential target event.

In Chapter 5, a follow-up study was conducted to investigate whether people predominately organize photographs by events or by special themed categories. As part of this study we also assessed the composition of the annotations associated with these picture groupings. Although, the results that were presented in this chapter may not be generalizeable to the population at large due to the small and selective sample, they are indicative of practice among college students who use the windows explorer to manage their photographs and are useful for design within this population group. The results from this study also provide corroboratory evidence to support some of the observations in Chapter 4. Based on observations, semi-formal interviews and a thorough inspection of each participants photo collection we found that participants would sort their photographs by events when importing them into the photo collection. Photographs from other people that co-experienced the event were merged with the user's own photographs from the event. We found evidence showing that some people organize photographs into special themed categories. However, even these participants would organize by events and would then place each event into an appropriate category. We found no evidence showing that pictures are sorted directly into categories. Although, we found some non-event related categories, such as "scenic pictures", these were usually created as part of a search task where the pictures were obtained from events folders. There were significantly more event groups than special themed categories. The fact that all the participants organized photographs by events seems to support an earlier observation in Chapter 4, where it was noted that people tend to think of their photographs in terms of events. In addition to this, we also found that on average 4.3 (s.d. 0.9) different types of information (*who*, *when*, *where*, *why* and *what*) were encoded in a photo collection. We found that the *who*, *where*, *when*, and *what* categories were used significantly more than the *why* category. The *where* category was used significantly more than the *who* category. There were no other significant differences between the different types of information. The fact that different types of information were used was more evidence for the need to support multiple search methods to allow users to make use of any information that is available to them.

The findings in Chapter 6 validate the some of the observations that were made in the study in Chapter 4. One hypothesis was that as the need to locate events is central to the three searching tasks, we expected to see the same search methods being used to complete each of the three common searching tasks. We found no significant difference in the usage of any of the six search methods across the task types. Furthermore, the search method usage was similar for each task (using the Kruskal-Wallis test at $p=0.05$). We also found that multiple search methods were used. On average 5.6 (s.d. 2.3) search strategies were used by each participant. In fact, users were quite creative in combining search methods to achieve their goals. Ten out of the twelve participants used one or more unique search sequences. In

Chapter 4, we found that one the factors that affected search strategies was memory (how much information was known about at event). In Chapter 6, we found that user goals were quite varied across knowledge categories (*Precisely-known*, *Something-known* and *Unknown*). In the *Precisely-known* category, we expected most of the search strategies to revolve around the *Keyword search* as we thought this would provide the quickest access to an event. However, this was not always the case as some participants felt that it required less effort to scroll through the list of folders or use the *Timeline browser* than bringing up the keyboard and typing a search string. In this category, participants would use the information at hand to directly access a target event using the least mentally, physically and temporally demanding search method. In the *Something-known* category, the participants would first reduce search space. The first strategy was to use the *Timeline browser* to narrow the search down to particular time period. Participants would typically begin by picking the year, then the time of year and would then narrow the search down to a particular month. The second strategy was to cluster similar events by searching for a common keyword. The third search strategy was to search for a pre-existing category. This essentially restricts the search space to the category. The fourth strategy was to think of a small set of potential targets and then to work through them until the target folder was found. In the *Unknown* category, the participants would quickly refresh their memory by quickly skimming through relevant portions of the photo collection before beginning a more focused search. The specific strategies that were used varied from skimming through the folders to jumping to various temporal locations.

Based on the search strategies that were recorded in each of the three knowledge categories, we proposed some guidelines for designing photo search tools. When events are well-known, the search strategy is to minimize the time and effort required to retrieve them. The goal is to minimize both. One way of doing so is by designing a set of highly specialized and fined tuned search techniques that allow people to use the information at hand to directly access an event. As events are less well-known, the search strategy is to reduce the search space and then look for target events within the reduced set. Integrating multiple search methods allows users to input as much information as they can to narrow the search space. When events are unknown the search strategies are more explorative to support learning and discovery. For this category, the goal of the search strategy is to maximize the knowledge gained, while minimizing the amount of interaction. On way of doing so is by allowing users to categorize events based on semantic properties. For example, with our search tool users were able to search for a common keyword to group semantically similar events. We also found some contributing factors that influence search strategies: context availability, user expertise, cognitive styles (the habitual and preferred way of doing cognitive tasks) and cognitive abilities (intellectual aptitude at performing particular cognitive tasks). In Chapter 6, we discuss the potential impact these factors have on design.

7.2.2 Photo search interface

The main contribution was the design of a single photo search interface that integrates multiple search methods. The design goal was to provide rapid access to events. The blend of different search methods was chosen to support users when the information about an event was well-known (e.g. using a *Keyword search*), less well known (e.g. using the *Timeline filter* to narrow the search space) or unknown (e.g. using the *Hierarchical folder-based browser* to refresh the users memory). Users were able to use multiple search methods when performing a task. Any constraint that was applied on one search technique was immediately reflected on the other techniques (e.g. When the timeline filter was applied, subsequent keyword searches were restricted to the specified time frame). This allowed users to input as much information as they could to narrow the search space.

Other contributions were the visual photo search techniques that were also integrated into the photo search interface. In contrast with previous work in this area, the new techniques support user's needs more adequately by: allowing the presentation mode to be dynamically configurable; providing a control modality that dynamically adjusts the presentation rate; supporting personalization; and making optimal use of the human visual system (by maintaining the optimal rate at which users can process visual information). Unlike previous studies, the two techniques were compared against a traditional thumbnail-grid browser for their ability to perform three common searching tasks, locating *events*, *singles* and *properties*. Results from the 72 participant study, confirm that the new techniques offer a substantial improvements over current approaches. The new techniques support faster navigation to target photographs and more accurate identification of photographs. The subjective workload is also lower for the new techniques. The *AutoZoom* technique was preferred for navigating through photographs as it reduces the motion blur when scrolling rapidly, provides context and has very simple controls. The lack of flexibility in controlling the scroll speed and zoom level was frustrating for some participants when inspecting photographs. The *ManualZoom* technique was preferred for inspecting photographs. The ability to zoom was used to gain context, as opposed to being used as a tool for controlling the visual flow of information. In fact, when moving rapidly to other locations participants did not mind the motion blur. Participants found the interactive and dynamic nature of both techniques fun and engaging. They also commented on the suitability of the techniques for serendipitous browsing. Both techniques provide high granularity controls. This makes them best suited for short to medium navigation distances. Using them to browse through large image collections for long periods of time can result in dizziness and eye fatigue. It can also be very time consuming. These techniques need to be coupled with other techniques that enable users jump to various locations in photo collection quickly and easily.

We also proposed algorithms for supporting the *AutoZoom* and *ManualZoom* techniques in limited environments. Previous research on SDAZ has only provided partial explanations of how to implement the technique, choosing to focus more on the interaction mechanism and its effect on users. Chapters 3 and 4 provide a clear implementation guide for developers and practitioners. The algorithms tackle three major limitations with mobile devices: limited screen size, different types of input mechanisms, and limited device capabilities. Algorithmic extensions were made to SDAZ to make it more suitable for small displays. A mapping function was used to support the different types of input mechanisms. A number of optimizations were presented on how to implement the techniques in processor and memory efficient manner, ensuring that they scale elegantly and are independent of the photo collection size. This is important because future devices will be able to store significantly larger photo collections.

7.3 Critical review of thesis

This section reviews the overall success of the user-centered design methodology and discusses the strengths and weakness of each of the user studies.

7.3.1. User-centered design methodology

Overall, this methodology was successful in generating guidelines, conceptual frameworks and models to guide the design of future systems. Through engaging in the process of design, novel artifacts were also created (such as the *AutoZoom* and *ManualZoom* techniques). The incremental nature of a user-centered design approach was beneficial because it allowed comparisons with current photo search tools. It also meant that the new techniques were able to be built within the resource constraints available.

However, a common criticism of an incremental approach is that the designs can sometimes amount to a little more than local optimizations [16]. In this research, this is certainly not that case as the incremental design approach was successful in generating a number of contributions. Through an iterative design process we were able to arrive at a radical design (see searching tool in Chapter 6 compared to the one in Chapter 3).

7.3.2 User studies

In Chapter 3 a usability study was conducted to compare the *AutoZoom* and *ManualZoom* interfaces against a more traditional interface. One limitation with the experiment was the fact that the study was simulated on computer by setting the viewport size to the same resolution as a HP Pocket PC (320x240) and using a mouse as a surrogate for a stylus. At the time of the study devices such as the HP Pocket PC did not have sufficient processing power and memory to run such applications. Although the study was simulated, the key factor for investigation was the screen size rather than the interaction device (hence the effort of modifying SDAZ for small displays). In fact a number of other studies in this area have conducted similar experiments [29][118]. In retrospect, replicating the experiment on a handheld device would have added very little to the overall thesis goal as the core goal of the experiment was merely to provide a foundation from which to ask more interesting questions about photo search.

A second concern was the fact personal photo collections were not used for each participant. A single set of 300 of photographs was used to provide a consistent set of stimuli across all tasks, subjects and conditions. This was done to ensure we could easily isolate the effects that were due to the other independent variables and also reduce the number of independent variables. Extensive training was also conducted to familiarize users with the set of photographs and to minimize learning effects when conducting the experiments.

The user studies in Chapters 4, 5 and 6 were conducted using personal photo collections on target devices, making them more ecologically valid. However, one limitation was the number participants that took part in the study as this limits the generalizability of the results. Another concern was the fact all the participants were drawn from a university environment. The concern here was that different types of users (e.g. teenagers or the elderly) taken from different environments (different socio-cultural factors at play) can have different searching behaviors and needs. The first important point to make is that this subject is contentious within the field of HCI and has been debated on numerous occasions [71][83][87][121][124]. While having a larger and more varied sample in Chapters 4, 5 and 6 would have increased the external validity of the results by making them more generalizable to the broader population, using 12 participants is an accepted compromise. Similar studies to the ones presented in Chapters 4 and 6 by Rodden and Wood [105] and Kirk et al. [62] have also used 12 participants. These papers are widely cited and have been accepted at CHI (the leading academic forum in the HCI community). In general, this criticism can be applied to a large body of work in this area. Regardless of this fact, Lindgaard and Chattratichart [71] contend that the question of "how many users" is the wrong way to think about it. For example, in our study we are looking for a mismatch between our model of how people search for photographs and the users' mental model on the key and critical tasks. Framed this way, the key criterion that determines the number of problems that are found is not the number of users they are, but rather how many tasks the participants try. Lindgaard and

Chattratichart [71] found that there was no reliable correlation between the number of participants tested and the number of usability problems uncovered. However, they found a significant correlation between the number of tasks that are evaluated and the number of problems that are uncovered. They conclude that with all other things equal, the better predictor for the number of usability problems that are uncovered is the number of tasks that users conduct in the study and not the number of participants that take part in the study. In Chapter 4 great care was taken to create a broad and balanced range of tasks: searching for small and large events at short and long navigational distances; searching for photographs with small and large features at short and long navigational distance; and searching for groups of photographs with different themes (or properties). Each participant conducted a total of 27 tasks. In Chapter 6, the questions for the three common search tasks were based on four themes, precise, vague, verificative and exploratory. Each participant was required to conduct a total of 12 tasks.

Although, the results presented in Chapter 5 may not be generalizeable to the population at large due to the small and selective sample, they are indicative of practice among college students who use the Windows Explorer to manage their photographs and are therefore useful for design within this population group. In fact, some statistics on the recent social networking phenomenon, Facebook, show why it is important to consider this population group. According to a study by Emergence Media, 50% of all US users are college students, 21% are high school students and 29% are undeclared [33]. Furthermore, most Facebook users are between the ages of 18-24 [33]. We focus on Windows users as Windows usage accounts for 87.2% of total operating system usage, while Mac only accounts for 3.9% (and Linux 3.3%) [134]. The main of objective of this study was to provide corroboratory evidence for some of the observations in Chapter 4 to assist us in formulating a set of more empirically grounded requirements to feed into the next iteration in the design process that is presented in Chapter 6. The results obtained in this study were adequate for this more tightly focused scope.

7.4 Future Work

To further validate the results in Chapter 3, one extension would be to replicate the experiment on target devices. A more ecologically valid experiment would involve using personal photographs and a variety of tasks (precise, vague, verificative, and explorative). A within-subject design would have to be used to allow each participant to be exposed to all the variables. However, great care would have to be taken to minimize the workload for participants. For example, to replicate the experiment in Chapter 3 using a within-subject design, each participant would have to perform 81 tasks.

In Chapter 4, the *AutoZoom* technique was preferred for navigating through photographs as it reduces the motion blur when scrolling rapidly, provides context and has very simple controls. The lack of flexibility in controlling the scroll speed and zoom level was frustrating for some participants when inspecting photographs. The *ManualZoom* technique was preferred for inspecting photographs. However, a few participants found it difficult to use the two controls. Even though, the *ManualZoom* technique was used more than *AutoZoom* technique in Chapter 6, both techniques have there respective strengths and weakness. Future work should investigate other visual photo search techniques that combine the strengths of both these techniques. In terms of supporting these techniques in limited environments, further research is also needed to find the optimal way to load thumbnails.

In Chapter 5, all the users in our dataset accumulated photographs from other people such as friends, family or colleagues. In our dataset, we found that on average 32% (s.d. 22.96) of all folders in a photo collection contain photographs from other people. This result is particularly interesting as it relates to the device synchronization problem that was observed in Chapter 4, where photographs from other people are placed in the wrong chronological order when cameras are not synchronized by date and time. Although the actual figure (i.e. 32%) might be higher than expected (due to the sample comprising mainly of university students), the significance of this result is that it highlights the synchronization problem and the need for adequate solutions to address to problem, particularly for groups of people that share a large proportion of the photographs that they capture. Further research is needed to find ways of synchronizing the timestamps on sets of images from an event that are captured using different devices. One way of addressing this issue would be to compare different sets of images from an event based on visual properties, taking into consideration the relative time difference between photographs. The sets of images can then be merged together based on best fit. The timestamps can then be readjusted based on a set of images that are correctly dated. However, this problem is more challenging when the temporal data does not exist (e.g. when the photographs are scanned).

In Chapter 6, we found that some participants grouped semantically similar events by searching for common keywords that were used to describe events. If users are more consistent in their labeling conventions it can open up new ways for locating photographs based on the type of information (i.e. who, when, where, what and why) and the nature of the information within these categories. Annotation and searching tools need to be more tightly coupled so that the benefits of annotation are more apparent to the user. The challenge here is to find a way of demonstrating the value of being more consistent in labeling conventions. One way is by showing compelling new ways of searching when photographs are imported into a photo collection. The solution must take into account the fact that users are willing to group photographs by events and provide a description, but are not willing to annotate each individual photograph.

To build on the research presented in this thesis, further research should focus on developing solutions that address the three knowledge categories: *Precisely-known*, *Something-known* and *Unknown*. Chapter 6 highlights the requirements for each of these categories and also provides examples of each. A single configuration (or bundle) of search techniques is unlikely to suit all users. For example, in our study we found that when participants were locating events, some would think of their events in terms of time, while other would think of them in terms of their pre-created categories. Temporally-oriented users prefer events to be ordered chronologically, while categorically-oriented participants prefer an alphabetical listing. Choosing one over the other will obviously impact negatively on some participants. One solution is to allow users to choose the techniques that should be integrated together to cater for their individual needs. The key to supporting customization is to ensure that each technique is built around locating events. This makes it easier add or remove a search technique. Further research is needed to look at how tools can be integrated and customized around possible "spaces of interaction" (For example in Chapter 6, multiple search methods were integrated around events). Another challenge is how to manage the transition as user's bridge the knowledge gap. For example, the techniques that allow users to explore and learn will differ from techniques that allow users to reduce the search space, which in turn will differ from the techniques that allow user to access the target directly. Our solution presents multiple techniques in a single user interface so that they are all accessible. This can impact the usability of the search techniques as they are all required to share the limited screen real estate. To make more efficient use of the limited screen space, a better solution would be to only show the techniques that are currently being used. The challenge then becomes how to manage the transition from one technique to the next.

In Section 7.3.2, it was noted that socio-cultural factors can possibly influence searching behavior and result in different user needs. Studying the social aspect of photo searching begins with the recognition that users are 'driven not simply by a set of internal goals and cognitive processes, but by the social setting in which people find themselves, by the action of others, individually and collectively, and by the social nature of the work being conducted and the goals sought' [85]. Further research is needed to explain how socio-cultural factors can influence the strategies, tactics and behavioral patterns followed by users. In a social setting, photo searching might be done collaboratively, especially when searching through archives that record a group's history. Further research is needed on how to facilitate this process, as well as other over-arching activities such as co-present photo sharing.

Appendix A

Experiment 1

This appendix contains the materials for the experiment described in Chapter 3.

A.1 Experiment Description

Thank you for agreeing to participate in this experiment. The basic aim of this experiment is to test a new photo browser for handheld devices. You will use the photo browser to search through a set of 300 photographs from a recent trip to New Zealand. You will go through a training session to prepare you for the experiment. This is to make sure that you are familiar with the photo collection, the photo browser and the tasks that will be used in the experiment. A description of the photo browser and the tasks is provided. The experiment director will demonstrate the photo browser and the tasks to you. The experiment director is there to help you, so please feel free to ask any questions. For the main experiment, you are required to complete 27 tasks. **The goal is to complete each task as quickly as possible**. However, you can control the pace of the experiment by selecting when you are ready to move on to the next task. This allows you to have a break between tasks. Here is a breakdown of everything you need to do:

1. Fill out a consent form
2. Fill out Part A of the questionnaire.
3. Training 1: Familiarize yourself with the photo collection.
4. Training 2: Read a description of the photo browser and practice using it.
5. Training 3: Read a description of tasks and practice performing them.
6. Take a short break.
7. Begin the main experiment.
8. After the main experiment, fill out Part B of the questionnaire.

A.2 Consent form

I, ... , agree to participate in testing a photo browser for handheld devices for the purposes of research and publication.

In giving my consent I acknowledge that:

1. The procedures required for the experiment have been explained to me.

2. I have been given an opportunity to ask questions about the experiment and the questions have been answered to my satisfaction.

3. I consent to the collection of data on my usage of the system and any other data relating to my behavior, experiences, perceptions, and opinions.

4. I understand that I am free to withdraw from the study at any time, without incurring any penalty or affecting my treatment or my relationships with the researcher now or in the future.

5. I understand that no information about me will be used in any way which reveals my identity.

Signed:.. Date: ....

A.3 Questionnaire - Part A

Please tick or circle where appropriate.

Name ______________________________
Age ______________________________
Gender ______________________________
Degree ______________________________
Majors ______________________________
Profession ______________________________
First Language ______________________________

1 How many hours a week do you spend playing computer games?

never	0	1-2	2-3	5-10	20+	**often**

2 Do you play games on your cell phone? Y N

3 Have you used a stylus-based interface before? Y N

4 Do you own a camera phone? Y N

5 Have you ever navigated a document by simultaneously scrolling and zooming? Y N

6 Which category most accurately describes your computer skills?

Novice user Can use office programs such as Word. Can also email and browse the web	Intermediate user Can use advanced applications such as graphics or web design programs	Expert user Computer programmer

7 Which category most accurately describes your photography skills?

Never taken a photo	Casual photographer	Expert photographer

A.4 Questionnaire - Part B

Consider the tasks that you have just completed. Please answer the following questions in relation to the tasks that you have just completed.

Place a circle around the **numeric** value that best reflects your response.

1. How much mental and perceptual activity was required (e.g. thinking, deciding, calculating, remembering, looking, searching, etc.)?

little	1	2	3	4	5	**a lot**

2. To what extent were the tasks mentally demanding?

very easy	1	2	3	4	5	**very difficult**

3. To what extent were the tasks simple or complex?

simple	1	2	3	4	5	**complex**

4. To what extent were the tasks exacting or forgiving?

exacting	1	2	3	4	5	**forgiving**

5. How much physical activity was required (e.g., selecting, dragging, scrolling, etc.)?

little	1	2	3	4	5	**a lot**

6. To what extent were the tasks physically demanding?

very easy	1	2	3	4	5	**very difficult**

7. How successful do you think you were in properly completing the tasks?

not at all	1	2	3	4	5	**very**

8. How satisfied are you with your performance in completing the tasks?

not at all	1	2	3	4	5	**very**

9. How hard did you have to work (mentally and physically) to accomplish your level of performance?

not at all	1	2	3	4	5	**very**

10. Overall, to what extent did you become frustrated whilst carrying out the tasks?

not at all	1	2	3	4	5	**very**

Please could you comment on your experience, for example: what did you like; what didn't liked or do you perhaps have some suggestions?

A.5 Thumbnail browser

The *Thumbnail* browser displays photographs in a flat scrollable list. The photographs are ordered chronologically. The most recent photograph is shown at the top of the list and the oldest photograph is shown at the bottom of the list. A vertical scrollbar allows you to scroll up and down the list of photographs. You can enlarge a thumbnail by clicking on it. The photograph is smoothly animated in to fill the screen. To return to the thumbnail view, simply click on the enlarged photograph. Don't worry if this description doesn't make sense. The experiment director will demonstrate all this to you now. Feel free to ask any questions.

A.6 AutoZoom browser

The *AutoZoom* browser displays photographs in vertical list with the most recent photograph at the top and the oldest photograph at the bottom. To scroll through the photographs, click in the center of the screen and then drag away from the center of the screen. The drag distance is used to control the scroll speed. The browser automatically zooms out as you scroll. The faster the scroll speed the further it zooms out. The scroll direction can be controlled by dragging above or below the center of the screen. If the mouse button is released during a dragging action, the images are smoothly animated in until they are shown at their full size. You can also use the scrollbar if you wish. Don't worry if this description doesn't make sense. The experiment director will demonstrate all this to you now. Feel free to ask any questions.

A.7 ManualZoom browser

The *ManualZoom* browser displays photographs in vertical list with the recent photograph at the top and the oldest photograph at the bottom. To scroll through the photographs, click in the center of the screen and then drag vertically away from the center of the screen. The drag distance is used to control the scroll speed. The scroll direction can be controlled by dragging above or below the center of the screen. The zoom level can be controlled in a similar way by dragging horizontally away from the center of the screen. The larger the horizontal drag distance the more the view is zoomed out. The scroll speed and the zoom level can be controlled simultaneously by dragging diagonally. If the mouse button is release during a dragging action, the images are smoothly animated in until they are shown at their full size. You can also use the scrollbar if you wish. Don't worry if this description doesn't make sense. The experiment director will demonstrate all this to you now. Feel free to ask any questions.

A.8 Tasks

There are three types of tasks:

1. ***Locate a photograph***. The experiment software will show you the photograph to look for and will also describe it (e.g. find a picture of sky tower). The task can be completed by displaying the photograph on the screen and then clicking the 'finish task' button. The goal is to complete the task as quickly as possible.

2. ***Locate an event***. The experiment software will provide a description of the event (e.g. find a picture of the balloon festival). The task can be completed by displaying the event on the screen and then clicking the 'finish task' button. The goal is to complete the task as quickly as possible.

3. ***Locate a group of photographs that share a common property***. The experiment software will provide a description of a property (e.g. how many pictures have a boat in them). The task can be completed by entering the number the pictures and then clicking the 'finish task' button. **The goal is to be as quick and as accurate as possible.**

Appendix B

Experiment 2

This appendix contains the materials for the experiment described in Chapter 4.

B.1 Consent form

I, ... , agree to participate in testing a photo browser for handheld devices for the purposes of research and publication. I also authorize the usage of my personal photo collection for the purpose of this research project. In giving my consent I acknowledge that:

1. The procedures required for the experiment have been explained to me and any questions I have about the experiment have been answered to my satisfaction.

2. I consent to the collection of data on my usage of the system and any other data relating to my behavior, experiences, perceptions, and opinions. I understand that the data will be captured using video and audio recording.

3. I understand that no information about me will be used in any way which reveals my identity.

4. I understand that I am free to withdraw from the study at any time, without incurring any penalty or affecting my treatment or my relationships with the researcher now or in the future.

Signed:..   Date: ....

B.2 Questionnaire

Please tick or circle where appropriate.

Reference ______________________

Name ______________________

Age ______________________

Gender ______________________

Degree ______________________

Majors ______________________

Profession ______________________

First Language ______________________

1	Have you used a stylus-based interface before?	Y	N
2	Do you own a camera phone?	Y	N
3	Have you ever navigated a document by simultaneously scrolling and zooming? For example by using the Dynamic Zoom feature in Acrobat Reader.	Y	N

4 Which category most accurately describes your computer skills?

Novice user Can use office programs such as Word. Can also email and browse the web	Intermediate user Can use advanced applications such as graphics or web design programs	Expert user Computer programmer

5 Which category most accurately describes your photography skills?

Never taken a photo	Casual photographer	Expert photographer

B.3 Conceptual model extraction checklist

Basic interaction

Tap and hold anywhere on screen	
Drag from initial stylus position	
Changing the scrolling direction	
Understands how to scroll	
Understand how to zoom	
Understands end of collection interaction	
Can use tap and hold feature to control zoom level	

Control widget

Yellow dot	
Green arrow	
White arrow	
Green bar	
Crosshair	

Contextual information

Folder label	
Dates	
Color bands	
Collection position pointer	

B.4 Post experiment interview

The post experiment interview was very informal. The interview was loosely based on the questions below. Other exploratory questions were used to probe areas of interest more deeply (for example, to follow up any observations that were made during the experiment).

1. Can you describe the strategy that you used in this experiment to locate events, individual photographs and groups of photographs with a common theme?

2. How appropriate was the browsing technique in helping you with the tasks? Was there anything you liked or disliked in particular?

3. Were the folder labels, date information, color bands, photo collection position indicator useful in helping you with the task? Was there anything you liked or disliked in particular?

4. Did you notice that lower quality images were swapped in for high quality images when scrolling rapidly? Did this affect your ability to search in any way?

5. Were the images useful when they were displayed really small?

6. Do you have any suggestions or ideas on how to improve the techniques further?

Appendix C

Experiment 3

This appendix contains the materials for the experiment described in Chapter 5.

C.1 Consent form

I, ... , agree to participate in a study on photo annotation practices. I also authorize the usage of my personal photo collection for the purpose of this research project. In giving my consent I acknowledge that:

1. The procedures required for the experiment have been explained to me and any questions I have about the experiment have been answered to my satisfaction.

2. I consent to the collection of data relating to my behavior, experiences, perceptions, and opinions. I understand that the data will be captured using video and audio recording.

3. I understand that no information about me will be used in any way which reveals my identity.

4. I understand that I am free to withdraw from the study at any time, without incurring any penalty or affecting my treatment or my relationships with the researcher now or in the future.

Signed:.. Date: ....

C.2 Questionnaire

Please tick or circle where appropriate.

Reference ______________________________

Name ______________________________

Age ______________________________

Gender ______________________________

Degree ______________________________

Majors ______________________________

Profession ______________________________

First Language ______________________________

1 Which category most accurately describes your computer skills?

Novice user	Intermediate user	Expert user
Can use office programs such as Word. Can also email and browse the web	Can use advanced applications such as graphics or web design programs	Computer programmer

2 Which category most accurately describes your photography skills?

Never taken a photo	Casual photographer	Expert photographer

3 Can you please list all the cameras you have owned or have borrowed to take some pictures that are now in your photo collection.

____

____

____

____

____

C.3 Interview

1. Can you show me what you do when you import new photographs from your camera into your photo collection?

2. Do you use any particular strategy to organize photographs into groups of events?

3. Are your groups of photographs arranged any particular way, for example according to dates?

4. Do you organize your folders in to a hierarchy? Can we go through some examples showing how and why you arranged photographs in a hierarchy?

5. How do you go about labeling each group of photographs? Do you have a naming scheme? How do you decide what information should be attached to a group of photographs? Do you ever use dates when labeling your photographs?

6. Do you ever annotate individual photographs?

7. Do you use any photo management software to import photographs into your collection or to view them? If not, why not?

8. How do you organize photographs from other people?

9. Do you create any special categories or collections of pictures or events? How are these categories created? Do you ever add more pictures to them? Do you name or structure categories any differently?

10. Do you own more than one camera? If so, are there instances where you use one and not the other? Do you treat the images that are taken with different devices any differently?

Appendix D

Experiment 4

This appendix contains the materials for the experiment described in Chapter 4.

D.1 Consent form

I, ... , agree to participate in testing a photo browser for handheld devices for the purposes of research and publication. I also authorize the usage of my personal photo collection for the purpose of this research project. In giving my consent I acknowledge that:

1. The procedures required for the experiment have been explained to me and any questions I have about the experiment have been answered to my satisfaction.

2. I consent to the collection of data on my usage of the system and any other data relating to my behavior, experiences, perceptions, and opinions. I understand that the data will be captured using video and audio recording.

3. I understand that no information about me will be used in any way which reveals my identity.

4. I understand that I am free to withdraw from the study at any time, without incurring any penalty or affecting my treatment or my relationships with the researcher now or in the future.

Signed:.. Date: ....

D.2 Questionnaire

Please tick or circle where appropriate.

Reference ______________________________

Name ______________________________
Age ______________________________
Gender ______________________________
Degree ______________________________
Majors ______________________________
Profession ______________________________
First Language ______________________________

1 Have you used a stylus based interface before? Y N

2 Do you own a camera phone? Y N

3 Have you ever navigated a document by simultaneously scrolling and zooming? For example by using the Dynamic Zoom feature in Acrobat Reader. Y N

4 Which category most accurately describes your computer skills?

Novice user Can use office programs such as Word. Can also email and browse the web	Intermediate user Can use advanced applications such as graphics or web design programs	Expert user Computer programmer

5 Which category most accurately describes your photography skills?

Never taken a photo	Casual photographer	Expert photographer

D.3 Conceptual model extraction

1. The screen is divided in to 5 vertical zones, each with its own functionality. What do you think each zone does?

2. How can you constrain the interface to show photographs from this year? And the previous year?

3. Which months have photographs attached to them?

4. What do you think happens when you select the year or month tab?

5. How can you tell which folders are associated with a given month?

6. The folders are arranged in a vertical list. Have they been arranged in any particular order?

7. Select a folder that has subfolders. Can you see how they subfolders are arranged?

8. Select any folder. Which months are associated with the folder?

9. How can you access the oldest folder?

10. How would you expand and contract a folder?

11. When you select a folder, where is the most recent picture in the folder displayed?

12. How can you tell which pictures are associated with a folder?

13. What do you think the two icons along the bottom represent?

14. Tap on the keyboard icon. What do you think the other icons represent?

15. Search for any of your folders?

16. Minimize the search. How can you navigate through the search results?

17. What are the folders ranked according to?

18. What do you expect to happen when the search is closed?

19. Search for another folder. Only this time search for the second or third keyword in the folder name? What do you notice?

20. Select a 3 month period and perform a search. What do you notice?

21. How can you select an image? Select the image below. What do you notice?

22. Where can you find information on this photo? What does this information tell you?

23. What do the grey bands show you?

24. Select the *AutoZoom* technique. Tap and hold the stylus on the screen. Now drag away slowly. What do you notice?

25. What happens when you release the stylus from the screen?

26. What information does the control widget show?

27. Now release the stylus from the screen and then tap and hold? Repeat this again. What do you notice?

28. Now scroll through the entire collection.

29. What do notice is happening? How can you tell which folder you are in?

30. Now select the *ManualZoom* technique.

31. Tap and hold your stylus on the screen. Now drag vertically. What do you notice?

32. Now drag horizontally. What do you notice?

33. Drag diagonally. What do you notice?

34. Scroll through the entire collection.

35. Return to the overview mode.

D.4 Post-experiment interview

The post experiment interview was very informal. The interview was loosely based on the questions below. Other exploratory questions were used to probe areas of interest more deeply (for example, to follow up any observations that were made during the experiment).

1. Was the timeline useful as a navigational tool?
2. Was the timeline useful as a contextual marker?
3. Was the timeline filter useful?
4. Would you have liked any additional information to be displayed on the timeline?
5. Did you like the way the timeline was integrated with the folders? Is there any way the relationship between the two could have been portrayed any better?
6. Did you like the fact that your hierarchical folder structure is preserved in the interface? Is this something that you would expect?
7. Did you find the organization of all your folders in a single vertical list confusing?
8. Did you find the lines useful in showing the structure and relationship between folders?
9. Did you find the order in which your folders were arranged confusing?
10. Would you have liked to see more folders on the screen at once?

11. Do you think it was a good design choice to show the current event and photograph in the center of the screen? Was it useful to see pictures from the next event?

12. Was the number of pictures shown sufficient to obtain an overview of the event?

13. Did you find the search facility useful?

14. Did you like the ability to search for any keyword in the folder label?

15. Did you like the fact that the search was dynamic?

16. Did you prefer the *AutoZoom* or *ManualZoom* interface? Were there instances where you felt that one was better than the other?

17. Did you like the fact that the timeline, folders, keyword search and visual search techniques were all integrated seamlessly?

18. Which combinations were most useful and why?

19. Are there other techniques that you would have liked to be added or removed?

20. Having played with the application, do you think it would motivate you to do anything different in the way you organize or structure your collection?

21. Would you buy this application?

22. What did you like or dislike

23. Any ideas or suggestions on how to improvement the search tool?

Bibliography

[1] Abras, C., Maloney-Krichmar, D., Preece, J. (2004). User-Centered Design. In Bainbridge, William Sims (Ed.). (2004). *Berkshire Encyclopedia of Human-Computer Interaction*. Great Barrington, MA: Berkshire Publishing Group. ISBN: 9780974309125.

[2] Alexander, N. (2005, September 9). Personal Photo Annotation. *Masters Thesis*. University of East Anglia. Retrieved: December 24, 2007, from the University of East Anglia School of Computer Science Website: http://www.cmp.uea.ac.uk/Research/photos/Alexander_MSc2005.pdf.

[3] American Society if Indexers. (2005, October 14). *How Information Retrieval Started*. Retrieved: December 24, 2007, from the American Society of Indexers Website: http://www.asindexing.org/site/history.shtml.

[4] Apple iPod. (2007). *iPod Classic*. Retrieved: December 24, 2007, from the Apple Online Store Website: http://www.apple.com/ipodclassic/.

[5] Apple iPhoto. (2007). *Organize by Events*. Retrieved: 24 December 27, 2007, from the Apple Website: http://www.apple.com/ilife/iphoto/#events.

[6] Aris, A., Gemmell, J., and Lueder, R. (2004, December). Exploiting Location and Time for Photo Search and Storytelling in MyLifeBits. *Microsoft Research Technical Report*, MSR-TR-2004-12. Retrieved: December 24, 2007, from Microsoft Research Website: http://research.microsoft.com/research/pubs/view.aspx?type=Technical%20Report&id=807

[7] Bates, M. J. (1989). The design of browsing and berrypicking techniques for the online search interface. *Online Review*, 13(5), pp. 407-424.

[8] Bakker, N. H. (2001). *Spatial Orientation in Virtual Environments.* Deft University Press. ISBN: 978-90-407-2152-6.

[9] Beardsley, T. (1999, March). Humans Unite, Ben Shneiderman wants to make computers into more effective tools by banishing talk about machine intelligence. *Scientific American*, in Washington D.C., Retrieved: 24 December, 2007, from Scientific America Website: http://www.sciam.com/article.cfm?articleID=0009C382-6651-1C6D-84A9809EC588EF21&ref=sciam

[10] Beck E., Christiansen M., Kjeldskov J., Kolbe N. and Stage J. (2003). Experimental Evaluation of Techniques for Usability Testing of Mobile Systems in a Laboratory Setting. *In Proceedings of OzCHI'03*, in Brisbane, Australia, pp. 106-115.

[11] Bex, P., and Pawels, L. (1999, November). Social perspectives: posing for posterity. *Interactions* 6 (6), pp. 26-31.

[12] Boardman, R. (2004). Improving Tool Support for Personal Information Management. *Doctoral thesis.* Department of Electrical and Electronic Engineering. Imperial College. London. Retrieved: December 24, 2007, from Richard Boardman's Website: http://www.iis.ee.ic.ac.uk/~rick/thesis/boardman04-thesis.pdf

[13] Broadbent, D. E., and Broadbent, M. H. (1987). From detection to identification: Response to multiple targets in rapid serial visual presentation. *Perception and Psychophysics*, 42, pp. 105-113.

[14] Brooks, P. (1994). Adding value to usability testing, In Nielsen and Mack, *Usability Inspection Methods*, ISBN 0-471-01877-5, pp. 255–271.

[15] Card, S., Moran, T. P. & Newell, A. (1983). *The Psychology of Human-Computer Interaction.* Hillsdale, NJ, LEA, ISBN: 0898598591.

[16] Carroll, J. M. Making Use. (2000). *Scenario-Based Design of Human Computer Interactions*. MIT Press. ISBN: 0262032791.

[17] Carroll, J. M., Kellogg, W. A., Rosson, M. B. (1991). The Task-Artifact Cycle. In: Carroll, John M. (Eds.) *Designing Interaction: Psychology at the Human-Computer Interface*. Cambridge University Press. ISBN: 978-0521409216.

[18] Carroll, J., and Campbell, R. (1986). Softening Up Hard Science: Response to Newell and Card. *Human-Computer Interaction*, Vol. 2(3), pp. 227-249.

[19] Chiaramella, Y. (1997). Browsing and querying: two complementary approaches for multimedia information retrieval. *Hypermedia - Information Retrieval - Multimedia*, pp. 9-26.

[20] Chittaro L. (2006, March). Visualizing Information on Mobile Devices. *IEEE Computer*, 39 (3), pp. 40-45.

[21] Cockburn, A. and Savage, J. (2003). Comparing Speed-Dependent Automatic Zooming with Traditional Scroll, Pan and Zoom Methods. In People and Computers XVII, *Proceedings of the 2003 British Computer Society Conference on Human-Computer Interaction (BCS HCI 2003)*, in Bath, England, pp. 87--102.

[22] Cockburn, A., Savage, J., and Wallace, A. (2005, April). Tuning and testing scrolling interfaces that automatically zoom. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CH '05), in Portland, Oregon, USA, ACM, New York, NY, pp. 71-80.

[23] Creswell, J. W. (2003) *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*, 2nd ed. Sage Publications. ISBN: 978-0761924425.

[24] Cutting, D. R., Karger, D. R., Pedersen, J. O., and Tukey, J. W. (1992, June). Scatter/Gather: a cluster-based approach to browsing large document collections. In *Proceedings of the 15th Annual international ACM SIGIR*

Conference on Research and Development in information Retrieval (Copenhagen, Denmark, June 21 - 24, 1992). N. Belkin, P. Ingwersen, and A. M. Pejtersen, Eds. SIGIR '92. ACM, New York, NY, pp. 318-329.

[25] Czerwinski, M., van Dantzich, M., Robertson, G. and Hoffman, H. (1999, May). The Contribution of Thumbnail Image, Mouse-Over Text and Spatial Location Memory to Web Page Retrieval in 3D. In *Proceeding of Interact '99*, pp. 163-170.

[26] Davis, M., Smith, M., Stentiford, F., Bambidele, A., Canny, J., Good, N., King, S., and Janakiraman, R. (2006, January). Using Context and Similarity for Face and Location Identification. *In Proceedings of SPIE Internet Imaging VII*, VOL 6061, in San Jose, California. pp. 60610E.

[27] Davis, M., Van House, N., Towle, J., King, S., Ahern, S., Burgener, C., Perkel, D., Finn, M., Viswanathan, V., and Rothenberg, M. (2005). MMM2: Mobile Media Metadata for Media Sharing. *In: Extended Abstracts of the Conference on Human Factors in Computing Systems* (CHI 2005), in Portland, Oregon, ACM Press, pp. 1335-1338.

[28] De Bruijn, O., & Spence, R. Rapid serial visual presentation: a space-time trade-off in information presentation. (2000, May). *In Proceedings of the Working Conference on Advanced Visual Interfaces* (AVI 2000), in Palermo, ITALY, ACM, pp. 189-192.

[29] Derthick, M. (2002). Interfaces for Palmtop Image Search. *In Proceedings of the Joint ACM/IEEE Conference on Digital Libraries* (2002), pp. 340-341.

[30] Desurvire, H. W. (1994). Faster, cheaper. Are usability inspection methods as effective as empirical testing? In Nielsen, J., and Mack, R. L. (Eds.), *Usability Inspection Methods*, John Wiley & Sons, New York, pp. 173-202.

[31] Dix, A., Finlay, J., Abowd, G., and Beale, R. (1998). *Human Computer Interaction*. Second Edition, Prentice Hall. ISBN: 0-13-239864-8.

[32] Eckert, M. and Buchsbaum, G. (1993). The Significance of Eye Movements and Image Acceleration for Coding Television Image Sequences. In Watson, A. ed. *Digital Images and Human Vision*, MIT Press, ISBN: 978-0-262-23171-8, pp. 90-98.

[33] Emergence Media. (2007, November). *Facebook Demographic and User Statistics. Retrieved*: December 24, 2007, from the Emergence Media Website: http://www.emergence-media.com/2007/11/facebook-demographic-user-statistics-fun/.

[34] Farkas, L., Blom, J., and Uusitalo, S. (2006, April). Social Uses of Digital Imagery with a Focus on SmartPhones Applicability. *In Proceedings of CHI 2006 Workshop on Mobile Social Software*, in Montreal, Canada, pp 31-33.

[35] Flickr (2007). *Popular Tags.* Retrieved: December 24, 2007, from the Flickr Website: http://www.flickr.com/photos/tags/.

[36] Frohlich, D. and Tallyn, E. (1999, May). Audiophotography: practice and prospects. *In Proceedings of CHI '99 Extended Abstracts on Human Factors in Computing Systems*, in Pittsburgh, Pennsylvania. ACM, New York, NY, pp. 296-297.

[37] Gargi, U., Deng, Y., and Tretter, D.R. (2003). Managing and Searching Personal Photo Collections. *In Proceedings of SPIE Storage and Retrieval for Media Databases*, pp. 13-21.

[38] Gargi, U. (2003, August). Consumer Media Capture: Time-Based Analysis and Event Clustering. *HP Tech Report Number HPL-2003-165*, December 24, 2007, from the HP Technical Report Website: http://www.hpl.external.hp.com/techreports/2003/HPL-2003-165.pdf.

[39] Gentner, D. and Stevens, A. L. (1983). *Mental Models*. Hillsdale, NJ: Erlbaum. ISBN: 978-0898592429.

[40] Graham, A., Garcia-Molina, H., Paepcke, A., and Winograd, T. (2002, July). Time as essence for photo browsing through personal digital libraries. In *Proceedings of the 2nd ACM/IEEE-CS Joint Conference on Digital Libraries* (JCDL '02), in Portland, Oregon, USA, ACM, New York, NY, pp. 326-335.

[41] Gutwin, C. (2002, April). Improving focus targeting in interactive fisheye views. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems: Changing Our World, Changing Ourselves,* in Minneapolis, Minnesota, USA. CHI '02. ACM, New York, NY, USA, pp. 267-274.

[42] Halle, D. (1993). *Inside Culture.* Art and Class in the America Home. Chicago: University of Chicago Press. ISBN: 978-0226313689.

[43] Hanbury, A. (2006). *MUSCLE, Guide to annotation*, Version 2.12, Technical document. University of Vienna. Retrieved: 24 December, 2007, from the MUSCLE Website: http://muscle.prip.tuwien.ac.at/docs/metadata_spec.pdf

[44] Harada, S., Naaman, M., Song, Y., Wang, Q., and Paepcke, A. (2004, June). Lost in memories: Interacting with large photo collections on PDAs. In *Proceedings of Joint Conference on Digital Libraries* (JCDL '04), in Tucson, Arizona, USA, ACM press, pp. 325-333.

[45] Hart, S. & Staveland, L. (1998). Development of NASA-TLX (task load index): Results of empirical and theoretical research, in P. Hancock & N. Meshkati, eds, Human Mental Workload, *Elsevier Science*, pp. 139–183.

[46] Hassan-Montero, Y. and Herrero-Solana, V. (2006, October). Improving Tag-Clouds as a Visual Information Retrieval Interfaces. *In Proceedings of International Conference on Multidisciplinary Information Sciences and*

Technologies (InSciT '06), in Mérida, Spain. Retrieved: December 24, 2007, from Hassan's Website:
www.nosolousabilidad.com/hassan/improving_tagclouds.pdf

[47] Hewett, T., Baecker, R., Card, S., Carey, J., Gasen, M., Mantei, G., Perlman, G., Strong, G., Verplank, B. (2004), *Curricula for Human-Computer Interaction*, ACM Press, New York, NY, Retrieved: December 24, 2007, from ACM SIGCHI Website: www.acm.org/sigchi/cdg/cdg2.html

[48] Hix, D. and Hartson, H. R. (1993). *Developing user interfaces: Ensuring usability through product and process*. NY. John Wiley. ISBN: 0471578134.

[49] Igarashi, T. and Hinckley, K. (2000). Speed-dependent Automatic Zooming for Browsing Large Documents, *In Proceedings of the 2000 ACM Conference on User Interface Software and Technology (UIST 2000)*, in San Diego, CA, ACM Press, pp. 139-148.

[50] InfoTrend (2006, January). *Mobile Imaging Study*. Retrieved: December 24, 2007, from InfoTrend Website:
http://www.infotrends-rgi.com/home/Press/itPress/2006/1.18.2006.html

[51] Intraub, H. (1980). Presentation rate and the representation of briefy glimpsed pictures in memory. *Journal of Experimental Psychology: Human Learning and Memory*, 6 (1), pp. 1-12.

[52] Jacucci, G., Oulasvirta, A., Salovaara, A., and Sarvas, R. (2005, November). Supporting the shared experience of spectators through mobile group media. In *Proceedings of the 2005 international ACM SIGGROUP Conference on Supporting Group Work* (GROUP '05), In Sanibel Island, Florida, USA, ACM Press, New York, NY, pp. 207-216.

[53] Jaimes A. (2006, January). Human Factors in Image Retrieval. *In IS&T/SPIE Proceedings of SPIE Vol. 6061, Internet Imaging VII*, in San Jose, CA, USA, pp. 606103.

[54] John, B. E., and Kieras, D. E. (1996). The GOMS family of user interface analysis techniques: Comparison and contrast. *ACM Transactions on Computer-Human Interaction*, Vol. 3 (4), pp. 320-351.

[55] John, B. E., and Kieras, D.E. (1996). Using GOMS for user interface design and evaluation: Which technique? *ACM Transactions on Computer-Human Interaction*, 3(4), 287-319.

[56] Jones, M., and Marsden, G. (2006). *Mobile Interaction Design*. Wiley. ISBN 0-470-09089-8.

[57] Kang, H., and Shneiderman, B. (2000). Visualization Methods for Personal Photo Libraries: Browsing and searching in the PhotoFinder. *In Proceedings of the IEEE International Conference on Multimedia and Expo (ICME2000)*, pp. 15390–1542.

[58] Kang, H., and Shneiderman B. (2000, July). Direct Annotation: A Drag-and-Drop Strategy for Labeling Photos. In *Proceedings of the international Conference on information Visualisation*. IV. IEEE Computer Society, Washington, DC, pp. 88.

[59] Karat, C. (1994). Comparison of user interface evaluation methods. In Nielsen, J., and Mack, R.L. (Eds.), *Usability Inspection Methods*, John Wiley & Sons, New York, ISBN 0-471-01877-5. pp. 203-232.

[60] Kim, K.S. and Allen, B. (2002). Cognitive and task influences on web searching behaviour. *Journal of the American Society for Information Science*, 53(2), pp. 109-119.

[61] Kindberg, T., Spasojevic, M., Fleck, R., and Sellen, A. (2005, April). The Ubiquitous Camera: An In-Depth Study of Camera Phone Use. *IEEE Pervasive Computing* 4 (2), pp. 42-50.

[62] Kirk, D., Sellen, A., Rother, C., and Wood, K. (2006, April). Understanding photowork. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems,* in Montréal, Québec, Canada, R. Grinter, T. Rodden, P. Aoki, E. Cutrell, R. Jeffries, and G. Olson, Eds. CHI '06. ACM, New York, NY, pp. 761-770.

[63] Kjeldskov J. and Stage J. (2004). New Techniques for Usability Evaluation of Mobile Systems. *International Journal of Human Computer Studies* (IJHCS) Elsevier, 60, pp. 599-620.

[64] Kjeldskov, J. and Graham, C. (2003, September) A Review of MobileHCI Research Methods. In *Proceedings of the 5th International Conference on Mobile Human-Computer Interaction* (MobileHCI03), in Udine, Italy, Springer-Verlag, pp. 317-335.

[65] Khella, A., and Bederson, B. B. (2004, October). Pocket photoMesa: A Zoomable image browser for PDA. In *Proceedings of MUM'04*, pp. 27-29.

[66] Korn, F. (1995). A Taxonomy of Browsing Methods: Approaches to the 'Lost in Concept Space' Problem. *Technical Report*, University of Maryland, Dept. of Computer Science, Retrieved: December 24, 2007, from Human Computer Interaction Lab Website:
www.cs.umd.edu/hcil/visible-human/vhp/graphbrowse.ps

[67] Koskinen, I., Kurvinen, E., and Lehtonen, T. (2002). *The Mobile Image.* IT Press, ISBN: 978-9518266368.

[68] Koskinen, I., Kurvinen, E., and Lehtonen, T. (2002). Mobile Digital Pictures-the future of the postcard? Findings from an experimental study. Teoksessa

Laakso & Östman (toim) Postikortti suhteessa kulttuurisiin tapoihin / *Postcards and Cultural Rituals*. Korttien talo, Hämeenlinna, pp. 69-96.

[69] Krauzlis, R. and Stone, L. (1999). Tracking with the Mind's Eye. *Trends in Neurosciences*, 22. pp. 544-550.

[70] Krishnan, A., Interactive Time-Centered Workspace Visualization. (2003). *MCMS Thesis*, Department of Computer Science, University of Waikato, Hamilton, New Zealand. Call no. TK7882.I6K75.

[71] Lindgaard, G. and Chattratichart, J. (2007, May). Usability testing: what have we overlooked? In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '07), in San Jose, California, USA, ACM, New York, NY, pp. 1415-1424.

[72] Mack, R. L., and Montaniz, F. (1994). Observing, predicting and analyzing usability problems. In Nielsen, J., and Mack, R. L. (Eds.), *Usability Inspection Methods*, John Wiley & Sons, New York, ISBN 0-471-01877-5, pp. 293-336.

[73] Mackay, W. E. (1998). Triangulation within and across HCI disciplines. *Human-Computer Interaction*. Hillsdale, New Jersey, Lawrence Erlbaum Associates. Commentary on: "Damaged Merchandise? A Review of Experiments that Compare Usability Evaluation Methods", W.D Gray and M.C. Salzman. Vol. 13(3), pp. 310-315.

[74] Mäkelä, A., Giller, V., Tscheligi, M., and Sefelin, R. (2000, April). Joking, storytelling, artsharing, expressing affection: a field trial of how children and their social network communicate with digital images in leisure time. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '00)*, Hague, Netherlands. ACM, New York, pp. 548-555.

[75] Marchionini, G. M. (1995). *Information seeking in electronic environments*. Cambridge University Press. ISBN: 0521586747.

[76] Maurer, D. (2006, March). *Four Modes of Seeking Information and How to Design for them*. Retrieved: December 24, 2007, from Boxes and Arrows Website: http://www.boxesandarrows.com/view/four_modes_of_seeking_information_and_how_to_design_for_them.

[77] Marcel, A. J. (1993). Conscious and preconscious recognition of polysemous words: Locating the selective effects of prior verbal context. In: Nickerson R S (ed.) *Attention and Performance*. Erlbaum, Hillsdale, NJ, Vol. VII, pp. 435-457.

[78] Mazza, R. (2004, March). *Introduction to Information Visualization*. University of Lugano. Faculty of Communication Sciences. Retrieved: December 24, 2007, from Mazza's Website: www.inf.unisi.ch/assistants/mazza/infovis_introduction.pdf

[79] McGrath, J., Martin, J. and Kulka, J. (1982). *Judgment Calls in Research*. CA, Sage Publications, ISBN: 0803918747.

[80] Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *The Psychological Review*, 63, pp. 81-97.

[81] Morgan, M. and Benton, S. (1989). Motion-deblurring in human vision. *Nature*, 340, in London, pp. 385-386.

[82] Moellic, P., Hede, P., Grefenstette, G., and Millet, C. (2005). Evaluating content based image retrieval techniques with the one million images CLIC Testbed. In *Proceedings of the Second World Enformatika Congress* (WEC'05), pp. 171-174.

[83] Molich, R., Ede, M. R., Kaasgaard. K., and Karyukin, B. (2004). Comparative usability evaluation. *Behaviour and Information Technology*, 23 (1), Taylor & Francis (2004), pp. 65-74.

[84] Morrison, D. R. (1968, October). PATRICIA - Practical Algorithm to Retrieve Information Coded in Alphanumeric. *Journal of the ACM*, 15(4), pp. 514-534.

[85] Munro, A.J., Hook, K., and Benyon, D. (1999). *Social navigation of information space*. London: Springer Verlag. ISBN: 1-85233-090-2

[86] Naaman, M., Song, Y. J., Paepcke, A., and Garcia-Molina, H. (2004, June). Automatic organization for digital photographs with geographic coordinates. In *Proceedings of the 4th ACM/IEEE-CS Joint Conference on Digital Libraries* (JCDL '04), in Tuscon, AZ, USA, ACM, New York, NY, pp. 53-62.

[87] Nielsen, J. and Landauer, T. K. (1993, April). A mathematical model of the finding of usability problems. *In Proceedings of the INTERACT '93 and CHI '93 Conference on Human Factors in Computing Systems*, in Amsterdam, The Netherlands, ACM, New York, NY, pp. 206-213.

[88] Nielsen, J. (1993). *Usability Engineering*. Academic Press, ISBN: 0-12-518405-0.

[89] Norman, D. A. (2006). Why doing user observations first is wrong. *Interactions* 13(4), pp. 50.

[90] Norman, D. A. (1999). *The Invisible Computer: Why Good Products Can Fail, the Personal Computer Is So Complex, and Information Appliances Are the Solution*. Boston: MIT Press, ISBN 0-262-64041-4.

[91] Norman, D. A. (1988). *The Design of Everyday Things*. New York, Doubleday Publishing Group, ISBN: 9780262640374.

[92] Norman, D. A. (1987). Some observations on mental models. *In Human-Computer interaction: A Multidisciplinary Approach*, R. M. Baecker, Ed. Morgan Kaufmann Publishers, San Francisco, CA, pp. 241-244.

[93] Norman, D. A. (1986). Cognitive Engineering. In Norman D. A. and Draper S. W. *User-Centered System Design: New perspectives in Human-Computer Interaction*. Hillsdale, NJ, Lawrence Earlbaum Associates, pp. 31-61.

[94] Norman D. A. and Draper S. W. (1986). *User-Centered System Design: New perspectives in human-computer interaction*. Hillsdale, NJ, Lawrence Earlbaum Associates. ISBN: 0898598729.

[95] O'Hara, K., Sellen, A., and Bentley, R. (1999, May). Supporting memory for spatial location while reading from small displays. In *CHI '99 Extended Abstracts on Human Factors in Computing Systems*, in Pittsburgh, Pennsylvania, ACM, New York, NY, pp. 220-221.

[96] Palmquist, R.A. and Kim, K.S. (2000). Cognitive style and online database search experience as predictor of web search performance. *Journal of the American Society for Information Science*, 51(6), pp. 558-566.

[97] Pirhonen, A., Brewster, S., and Holguin, C. (2002, April). Gestural and audio metaphors as a means of control for mobile devices. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems: Changing Our World, Changing Ourselves* (CHI '02), in Minneapolis, Minnesota, USA,. ACM, New York, NY, pp. 291-298.

[98] Plaisant, C. (2004, May). The challenge of information visualization evaluation. In *Proceedings of the Working Conference on Advanced Visual interfaces* (AVI '04), in Gallipoli, Italy, ACM, New York, NY, pp. 109-116.

[99] Platt, J C., Czerwinski, M., and Field, B. (2003, December). PhotoTOC: Automatic Clustering for Browsing Personal Photographs. In *Proceedings of the 2003 Joint Conference of the Fourth Pacific Rim Conference on Multimedia*, Vol. 1, in Singapore, pp. 6-10.

[100] Preece, J., Rogers, Y. and Sharp, H. (2002). *Interaction Design: Beyond Human-Computer Interaction*, John Wiley & Sons, Inc. ISBN: 9780471492788.

[101] Raymond, J. E., Shapiro, K.L., and Arnell, K. M. (1992). Temporary suppression of visual processing in an RSVP task: An attentional blink? *Journal of Experimental Psychology*. Human Perception and Performance, 18, pp. 849-860.

[102] Rekimoto, J., Oba, H., and Ishizawa, T. (2003, April). SmartPad: a finger-sensing keypad for mobile interaction. In *CHI '03 Extended Abstracts on Human Factors in Computing Systems*, in Ft. Lauderdale, Florida, USA, ACM, New York, NY, pp. 850-851.

[103] Robbins, D. C., Cutrell, E., Sarin, R., and Horvitz, E. (2004, May). ZoneZoom: map navigation for smartphones with recursive view segmentation. In *Proceedings of the Working Conference on Advanced Visual interfaces* (AVI '04), in Gallipoli, Italy, ACM, New York, NY, pp. 231-234.

[104] Robertson, G., Czerwinski, M., Larson, K., Robbins, D. C., Thiel, D., and van Dantzich, M. (1998, November). Data mountain: using spatial memory for document management. In *Proceedings of the 11th Annual ACM Symposium on User interface Software and Technology* (UIST '98), in San Francisco, California, United States, ACM, New York, NY, pp. 153-162.

[105] Rodden, K. and Wood, K. R. (2003, April). How do people manage their digital photographs? *In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '03), In Ft. Lauderdale, Florida, USA, ACM, New York, NY, pp. 409-416.

[106] Rodden, K. (2001, October). Evaluating Similarity-Based Visualisations as Interfaces for Image Browsing, *PhD thesis and Technical Report 543*, UCAM-CL-TR-543, ISSN 1476-2986, University of Cambridge Computer Laboratory.

Retrieved: December 24, 2007, from the University of Cambridge, Computer Laboratory Technical Reports Website:
www.cl.cam.ac.uk/techreports/UCAM-CL-TR-543.pdf

[107] Rodden, K., Basalaj, W., Sinclair, D., and Wood, K. (2001, April). Does organisation by similarity assist image browsing? In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '01), in Seattle, Washington, United States. ACM, New York, NY, pp. 190-197.

[108] Rodden, K. (1999, April). How do people organise their photographs? *BCS IRSG 21st Annual Colloquium on Information Retrieval Research*, in Glasgow, Scotland, pp. 1-11.

[109] Rohs, M., and Roduner, C. (2005, May). Camera Phones with Pen Input as Annotation Devices. In *Proceedings of the Workshop on Pervasive Mobile Interaction Devices (PERMID) at PERVASIVE 2005*, in Munich, Germany, pp. 23-26.

[110] Rother, C., Bordeaux, L., Hamadi, Y., and Blake, A. (2006, July). AutoCollage. *ACM Transactions on Graphics (SIGGRAPH).* 25(3), pp. 847-852.

[111] Sarvas, R. (2005, September). User-centric Metadata for Mobile Photos. *Pervasive Image Capture and Sharing Workshop at Ubicomp 2005*, in Tokyo, Japan, Retrieved: December 24, 2007, from the Digital Content Communities Research Group Website:
http://pong.hiit.fi/dcc/papers/Sarvas_UserCentricMetadata_final.pdf

[112] Shneiderman, B. (2003) *Leonardo's Laptop*. The MIT Press, ISBN: 0262194767.

[113] Shneiderman, B., and Kang. H. (2000, July). Direct Annotation: A Drag-and-Drop Strategy for Labeling Photos. In *Proceedings of the international Conference on information Visualisation*, IV. IEEE Computer Society, Washington, DC, pp. 88-95.

[114] Shneiderman, B. (1986). *Designing the User Interface: Strategies for Effective Human-Computer Interaction.* Addison-Wesley Longman Publishing Co., Inc. ISBN: 0-201-16505-8.

[115] Shapiro, K.L., and Raymond, J.E. (1998). Temporal allocation of visual attention: Inhibition or interference? *In* D. Dagenbach & T.H Car (eds), *Inhibitory mechanisms in attention, memory and language. Boston.* Academic Press, pp. 151-188.

[116] Simmel, G. (1950). Sociability. An example of Pure or Formal Sociability. In Kurt, H and Wolff (ed), *The Sociology of George Simmel,* Glencoe, IL: Free Press, ISBN: 978-0029289204 , pp. 40-57.

[117] Snavely, N., Seitz, S. M., and Szeliski, R. (2006, August). Photo tourism: exploring photo collections in 3D. In *Proceedings of ACM Transactions on Graphics (SIGGRAPH'06),* 25(3), in Boston, Massachusetts, ACM, New York, NY, pp. 835-846.

[118] Spence, B., Witkowski, M., Fawcett, C., Craft, B., and de Bruijn, O. (2004, May). Image presentation in space and time: errors, preferences and eye-gaze activity. In *Proceedings of the Working Conference on Advanced Visual interfaces* (AVI '04), in Gallipoli, Italy, ACM, New York, NY, pp. 141-149.

[119] Spence, R. (2002, March). Rapid, serial and visual: a presentation technique with potential. *Information Visualization* 1 (1), pp. 13-19.

[120] Spence, R. (1999). Rapid, Serial and Visual: a presentation technique with potential. *IJHCS,* 51(5), pp. 919-945.

[121] Spool, J. and Schroeder, W. (2001, March). Testing web sites: five users is nowhere near enough. *In CHI '01 Extended Abstracts on Human Factors in Computing Systems* (CHI '01), in Seattle, Washington, ACM, New York, NY, pp. 285-286.

[122] Tse, T., Marchionini, G., Ding, W., Slaughter, L., and Komlodi, A. (1998, May). Dynamic key frame presentation techniques for augmented video processing. In *Proceedings of Advanced Visual Interfaces (AVI'98),* in L'Aquila, Italy, ACM, pp. 185-194.

[123] Van House, N., and Davis, M. (2005). The Social Life of Cameraphone Images. In *Proceedings of the Pervasive Image Capture and Sharing: New Social Practices and Implications for Technology Workshop* (PICS 2005), at the Seventh International Conference on Ubiquitous Computing (UbiComp 2005), in Tokyo, Japan, Retrieved: December 24, 2007, from Nancy Van House's Website: http://people.ischool.berkeley.edu/~vanhouse/Van%20House,%20Davis%20-20The%20Social%20Life%20of%20Cameraphone%20Images.pdf.

[124] Virzi, R. A. (1992, August). Refining the test phase of usability evaluation: how many subjects is enough? *Human Factors,* 34 (4), pp. 457-468.

[125] von Ahn, L. and Dabbish, L. 2004. Labeling images with a computer game. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (*CHI '04), in Vienna, Austria,. ACM, New York, NY, pp. 319-326.

[126] Wang, W. (2004). Human Navigation Performance Using 6 DOF Dynamic Viewpoint Tethering in Virtual Environments. *PhD Thesis.* University of Toronto. Retrieved: December 24, 2007, from Wenbi Wang's Website: http://vered.rose.utoronto.ca/people/wenbi/

[127] Wang, P., Hawk, W.B, and Tenopir, C. (2000, January). Users' interaction with World Wide web resources: an exploratory study using a holistic approach. *Information Processing and Management, 36(2)*, pp. 229-251.

[128] Wang, Q., Harada, S., Hsieh, T., and Paepcke, A. (2005). Visual Interface and Control Modality: An Experiment about Fast Photo Browsing on Mobile

Devices, In *Proceedings of IFIP INTERACT'05*: Human-Computer Interaction 2005, pp. 240-252.

[129] Wang, M., Xie, X., Ma, W., and Zhang, H. (2003, November). MobiPicture: browsing pictures on mobile devices. In *Proceedings of the Eleventh ACM international Conference on Multimedia*, in Berkeley, CA, USA (MULTIMEDIA '03), ACM, New York, NY, pp. 106-107.

[130] Wickens, C. D. Frames of Reference for Navigation. (1999). *In D. Gopher & A. Koriat (Eds.), Attention and Performance Xvii, Vol.1.* Cambridge, MA: MIT Press. pp. 113-144.

[131] Wilhelm, A., Takhteyev, Y., Sarvas, R., Van House, N., and Davis, M. (2004, April). Photo annotation on a camera phone. In *CHI '04 Extended Abstracts on Human Factors in Computing Systems*, in Vienna, Austria, ACM, New York, NY, pp. 1403-1406.

[132] Whittaker, S., Terveen, L., and Nardi, B. A. (2000). Let's stop pushing the envelope and start addressing it: a reference task agenda for HCI. *Human Computer Interaction*, 15, pp. 75–106.

[133] Whittaker, S. and Sidner, C. (1996, April). Email overload: exploring personal information management of email. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems: Common Ground*, in Vancouver, British Columbia, Canada, M. J. Tauber, Ed. CHI '96. ACM, New York, NY, pp. 276-283.

[134] W3Schools. (2007). *OS Platform Statistics.* Retrieved: December 24, 2007, from the W3School's Website: http://www.w3schools.com/browsers/browsers_os.asp

[135] Yee, K., Swearingen, K., Li, K., and Hearst, M. (2003, April). Faceted metadata for image search and browsing. In *Proceedings of the SIGCHI*

Conference on Human Factors in Computing Systems (CHI'03), in Ft. Lauderdale, Florida, USA, ACM, New York, NY, pp. 401-408.

[136] Zeki, S. (1993). *A Vision of the Brain*. Oxford. Blackwell Scientific Publications, ISBN: 978-0-632-03054-5.

[137] Zunch Worldwide Inc. (2006, June). Search Engine Usage Statistics. Retrieved: December 24, 2007, from the Zunch Worldwide inc. Website: http://seo.zunch.com/search_engine_usage_statistics.htm.